

Calmer waters ahead

A conviction, the passage of a bill, and the arrival of some new committee chairmen have markedly improved the immediate outlook for biomedical research in the United States.

The past few years have witnessed a rough patch for relations between the US National Institutes of Health (NIH) and its overseers in Congress. Stagnant funding has followed on from the largesse that doubled the biomedical agency's budget between 1998 and 2003. Political leaders have also publicly doubted whether the \$28.6-billion agency is making the most of its windfall — and have excoriated agency officials for their allegedly lax approach to financial conflicts among its scientists (see *Nature* **443**, 252–253; 2006).

But two events occurred last week that, between them, have the potential to mark a turning point for the world's largest research agency. On Friday, the only NIH researcher to face criminal charges over conflict-of-interest issues arising from a forthright investigation by the *Los Angeles Times*, subsequently followed up in Congress, pleaded guilty in a federal court (see page 803). His sentencing on 22 December, along with new, tighter agency rules, may begin to disperse the cloud that has been hanging over the agency as a result of the actions of a few dozen NIH researchers.

Perhaps more importantly, in the small hours of 9 December, the outgoing Congress — pushed along by Congressman Joe Barton (Republican, Texas) — passed a bill that will guide the management of the agency over the coming years (see page 796).

This 'authorization' bill sets caps that could permit further increases in the NIH budget. This is no guarantee that the money will actually be doled out by the appropriations committees in Congress. But it is, says NIH director Elias Zerhouni, "a renewed vote of confidence in the NIH and really a turning point". Zerhouni asserts that after years of wondering aloud about whether the NIH would deliver on its doubled budget, Congress has finally decided that the agency is

doing a good job — and that it needs help to do a better one.

There are other reasons why the National Institutes of Health Reform Act of 2006 is important. Once President Bush signs the bill, the NIH will have, for the first time in 13 years, a broad, explicit and up-to-date law governing its administrative structure. The bill lays out some significant requirements that may in effect buy the agency something it has been lacking on Capitol Hill: goodwill. In particular, the bill requires the NIH director to submit biennial reports to Congress, detailing the work of its 27 institutes and centres, and justifying their priorities. It is this kind of accountability that many in Congress felt has been missing since the doubling of the budget; delivering on it will help the agency re-establish its reputation there.

At the same time, the bill establishes a 'common fund' comprising up to 5% of the agency's total budget. This will be used to fund multi-institute research that otherwise wouldn't happen, helping the NIH keep up as biomedical science evolves at unprecedented speed.

For many biomedical researchers, funding levels represent the clearest indication of whether goodwill has been re-established. In this regard, the omens are good, particularly given the allocation of critical committee positions in the new Congress. David Obey (Democrat, Wisconsin), a long-time champion of the NIH, will be chairing the relevant subcommittee in the House of Representatives, as well as the full appropriations committee. On the Senate side, Tom Harkin (Democrat, Iowa), another long-standing NIH devotee, is set to chair the subcommittee funding the agency.

Given this sort of backing, and last week's events, it is beginning to look as if the NIH could enjoy a happier and more prosperous new year than might have been envisaged just a few months ago. ■

An open debate

Researchers who work with animals should join the discussion on animal experimentation.

Animal research saves lives. That is the mantra often used to counter verbal and physical attacks on animal researchers and their institutions by animal-rights activists. And it is unquestionably true: animal research has made many valuable contributions to medical science.

However, the simplicity of the slogan barely does justice to the complexity of the issue. From a scientific point of view, for example, it is clear that certain animal models are useful: the neural prosthetics that promise to restore some independence to paraplegics, for example, arose from curiosity-driven studies of the primate brain (see *Nature* **443**, 122; 2006). But others are imperfect: certain mouse models of cancer, for example, do not accurately mimic the disease

in humans, and may even have hampered the development of some drugs (see *Nature* **442**, 739–741; 2006).

In this week's issue, several articles explore what scientists really think about animal research, and what the future may hold for such work. As part of our investigation, we surveyed many of our readers in the life sciences anonymously, to solicit their views on aspects of the topic. The 1,700 or so readers who responded to our online survey are not necessarily representative of the entire community of biologists, but their responses nonetheless offer some valuable insights into the views of working scientists on questions related to animal research. The complete survey results are published on the web at www.nature.com/news/specials/animalresearch. Given the reluctance of many scientists to speak on-the-record to our reporters on this particular subject (see page 808), the exercise has been helpful in generating an overall picture of how scientists themselves view this highly contentious public issue.

Three-quarters of respondents said that animal research was 'essential' to the progress of biomedical science. However, a minority



(fewer than one-fifth) express some misgivings about their work. “As a researcher in the field of HIV vaccine development, I am placed in a very awkward position regarding the use of non-human primates,” said one immunologist. “I personally feel uncomfortable with primate research yet I realize that without primate data, vaccine candidates are rarely forwarded to human trial.”

Fear is clearly a significant factor in excluding the voices of ordinary researchers from public discussion of these issues. No scientist should have to risk life and limb in order to speak about perfectly lawful work. Welcome steps have been taken by governments in both Britain and the United States to pass laws that will protect scientists under the law. And there is some evidence, in both countries, that public support for animal research has actually increased in response to animal-rights extremism.

Our survey also suggests that research agencies, universities and

other institutions could do more to ensure that scientists feel free to talk about their work. There are also some indications that peer pressure is not always conducive to open communication with the public about animal research. “I am more concerned that the scientific community, rather than the animal-rights movement, makes it difficult to voice a nuanced opinion on animal research,” said one neuroscientist, whose research involves using imaging to study the brains of human patients.

It is essential that researchers feel free to speak out, both within the community and beyond it, on this issue. If, as seems to be the case, scientists have made some headway in persuading the public of the value of animal research, then this is an opportune moment for them to engage in a full and open debate about the options that lie ahead — including improvements to research practice and the development of alternative approaches. ■

The elephant in the room

A biotech trial that went awry.

Sooner or later during the life of any drug candidate, there comes the time when it must be administered to humans. Phase I clinical trials, or ‘first in man’ trials, will always carry an inherent risk; the critical thing is how to manage the risk to the paid participants who volunteer as test subjects.

This is the question that a panel convened in Britain has sought to answer in the wake of the trial of TGN1412 — an antibody aimed at treating autoimmune diseases such as rheumatoid arthritis — that put six men in intensive care at London’s Northwick Park Hospital in March this year (see *Nature* **440**, 388–389; 2006). The panel was brought together to advise the Medicines and Healthcare Products Regulatory Agency (MHRA), which governs British clinical trials, on how to minimize the risks of such an event recurring.

Implicit in the list of 22 recommendations released by the panel on 7 December is an acceptance of the inherent dangers posed by some drug candidates — particularly those that target the immune system. The group’s final report describes such compounds as “high-risk drugs”, but makes no recommendations about when the MHRA should refuse to authorize a trial on the basis of unknown risk. Instead, the report leans towards learning as much as possible about such molecules before they are administered to people.

The panel recommends, for example, that the MHRA should consult external experts with specific knowledge of the area in question, such as a particular group of compounds. It also advocates creating a database of animal data, which might flag up adverse reactions observed previously that could otherwise be missed by regulators.

The proposals depart from the British regulator’s prior practice of approving clinical trials through a process that amounted to ticking a series of boxes, towards a more subjective consideration of how drug candidates actually work. It will no longer be enough simply to demonstrate that a compound is non-toxic in an animal model.

This is to be applauded. But the fact remains that many of the drugs now being readied for clinical trials, for all their enticing potential

benefits, carry risks that can never be eliminated. TGN1412, for example, was intended to bypass one of the cellular checks blocking the activation of a certain class of T cell. The problem was that, in humans, its activation was more widespread than expected among the different types of T cell, inducing a storm of inflammatory molecules that led to widespread inflammation and organ failure.

Preclinical trials in monkeys had shown no such adverse reaction, despite the fact that they share an almost identical receptor for the antibody. As part of the investigation that led to the report, immunologist Stephen Inglis of the National Institute for Biological Standards and Control near London and his colleagues developed an *in vitro* test using human blood cells that recreated the behaviour of the trial volunteers’ cells.

Inglis and the rest of the advisory panel now advocate including such tests in future preclinical evaluations of antibody drug candidates. But this test was retrospective — a team of immunologists took months to recreate something that bitter experience had shown to be possible. Using similar methods to spot dangerous compounds before they go to trial will always be an exercise in guesswork.

The panel’s advice included several recommendations that should be simple common sense. Volunteers should be dosed one by one, rather than simultaneously; doctors performing the tests should have specialist, rather than general, expertise; and trials should be carried out in dedicated centres within existing hospitals, with 24-hour emergency facilities.

All this represents necessary progress. Less satisfactory is the government’s failure so far to hold anybody accountable for the Northwick Park incident. The MHRA, in a subsequent investigation, exonerated both itself and Parxel, the company that administered the trial; TeGenero, the German firm that developed the drug, has gone bust. The victims have been appallingly maimed, suffering from amputations and with bleak health prospects. They are well represented legally and will doubtless now look to the courts, rather than the regulator, to determine where responsibility for their condition lies. ■

“It will no longer be enough simply to demonstrate that a compound is non-toxic in an animal model.”

RESEARCH HIGHLIGHTS

Horny beetles

Am. Nat. **168**, 711–729 (2006)

The function of horns on *Onthophagus* beetles (pictured) isn't just for combat or sexual prowess, but also so that the animals can break out of their larval shells.

Armin Moczek of Indiana University in Bloomington used electric currents to destroy the horn tissue of beetle larvae. When the altered larvae tried to hatch they were unable to break free of their shell. Moczek's team also discovered that all larvae have the horns regardless of sex — some females reabsorb their horns before reaching adulthood.

It remains unclear whether the evolution of the beetles' thick shell was made possible by the horns, or whether the need for a harder shell drove the appearance of stronger horns.



A. COOKE

SYNTHETIC CHEMISTRY

Gold road to safe sushi

Angew. Chem. Int. Edn doi:10.1002/anie.200601963 (2006)

Use of a new kind of chemical reaction catalysed by a gold compound has helped Craig Forsyth and his team at the University of Minnesota, Minneapolis, make an important segment of azaspiracid, a toxic marine molecule. This toxin accumulates in shellfish and when ingested by humans attacks the liver, pancreas and spleen. It may also be associated with neurological problems and tumour growth.

Forsyth's team achieved its synthesis by replacing the ketal group often used in similar reactions with an alkyne group. This prevented a carbon–carbon double bond within the molecule from shifting to a new position, a process that can disrupt the reaction. The method for making part of the complex azaspiracid molecule will aid the lab-based synthesis of related toxins, boosting attempts to study the molecules and treat their effects.

PALAEONTOLOGY

Unscrambled eggs

Geology **34**, 1037–1040 (2006)

Silty shales from China contain 500-million-year-old eggs preserved in a way never before seen in the fossil record.

A team led by Jih-Pai Lin of Ohio State University in Columbus used X-ray imaging to study deposits from the Kaili Formation in Guizhou province. The analysis revealed clusters of eggs dating from the Middle Cambrian, between 501 million and 513 million years ago, preserved three-dimensionally in silica. The team was even

able to distinguish cells in the act of dividing.

The discovery sheds light on the life histories of marine invertebrates that produced the eggs. It also raises the possibility of finding fossil embryos in other similar settings.

EVOLUTION

Old gene learns new tricks

Proc. Natl Acad. Sci. USA **103**, 19407–19412 (2006)

Seahorses and pipefish are unique among vertebrates in that pregnancy is a burden shouldered by males. This phenomenon may have evolved because dads co-opted an ancestral gene and used it to nurture embryos.

April Harlin-Cognato of Texas A&M University in College Station and her colleagues searched for genes that are highly expressed in the brood pouch (pictured below) — where eggs are fertilized and embryos grow — of pregnant male gulf pipefish (*Syngnathus scovelli*). They show that an ancient version of one of these genes may have worked in the liver and kidneys of bony fish.

The gene, which may have been drafted into the brood pouch more than 50 million years ago and has been dubbed *patristacin* by the authors, could help to control embryo ion content or help embryos to hatch.



GEOLOGY

Lava locks up carbon

J. Geophys. Res. **111**, B12201 (2006)

Flood basalts, giant outpourings of lava on Earth's surface, could be a good place to sequester carbon dioxide emitted by power plants.

The basalts come in layers with porous tops that could soak up carbon dioxide, and in areas — such as India and the US southeast — where coal-fired power plants are common. But little technical work has been done on the feasibility of sequestration in flood basalts, says a team led by Peter McGrail of the Pacific Northwest National Laboratory in Richland, Washington.

The team's lab experiments show that water saturated with carbon dioxide reacts with basalt to form stable minerals, adding weight to the idea that carbon dioxide pumped into basalt could be stably locked away.

GENETICS

Malaria parasite uncovered

Nature Genet. doi:10.1038/ng1924; 10.1038/ng1930; 10.1038/ng1931 (2006)

Geneticists have unveiled a slew of information on *Plasmodium falciparum*, the parasite that causes malaria. Three papers in *Nature Genetics* compare and contrast genome sequence data from a variety of strains — some that are used in the lab, some taken from infected people, and some from a species that infects chimpanzees.

By comparing the genomes, the researchers have identified which life stages of the parasite are more likely to show variation. They can also pick out more constant parts of the genome, which could serve as targets for treatments.

C. PARTRIDGE

BACTERIOLOGY

Fight on two fronts

Proc. Natl Acad. Sci. USA **103**, 18499–18503 (2006)

One way in which mammals infected with anthrax fight back is by trying to starve the bacteria of iron. Bacillibactin, a compound with which anthrax scavenges iron from the environment, is targeted by siderocalin, a part of the innate immune system.

Now Kenneth Raymond of the University of California, Berkeley, and his co-workers have shown how a second iron-scavenging molecule, petrobactin, allows anthrax to evade this control. The subunit with which petrobactin binds iron is very unusual, so siderocalin, which recognizes a range of iron-binding compounds, ignores this one.

Targeting petrobactin, the team suggests, might be a way of producing very specific anti-anthrax treatments.

GEOLOGY

Deep waters of life

Science doi:10.1126/science.1135013 (2006)

The arrival of oxygen in the world's oceans may have sparked the evolution of animal life.

Analysis of 580-million-year-old rock from the Avalon Peninsula in Newfoundland, Canada, suggests that oxygen entered the deep oceans at the end of an ice age known as the Gaskiers glaciation. Fossil evidence indicates that large multicellular organisms appeared on the sea floor shortly after this period of glaciation.

The rock data were generated by Don Canfield of the Nordic Center for Earth Evolution in Odense, Denmark, and his colleagues. Canfield's team studied levels of iron, sulphur and carbon in the rocks in order to characterize the chemical make-up of the oceans at the time when the rocks formed.

OCEANOGRAPHY

Southern warming

J. Clim. **19**, 6382–6390 (2006)

Global-warming predictions are currently hampered by uncertainties about the amount of heat and carbon dioxide that the Southern Ocean will take up. A computer model developed by Joellen Russell of the University of Arizona in Tucson and her colleagues now suggests that the ocean will be able to absorb more than previously thought.

The key is westerly winds that drive the powerful Antarctic circumpolar current. Russell's team locates these winds closer to the pole than in previous climate models, in line with recent observations. With the winds gradually moving poleward, the

ocean no longer stratifies into layers as rapidly, allowing heat and carbon dioxide to penetrate more deeply.

MATERIALS SCIENCE

Concrete evidence

J. Am. Ceram. Soc. **89**, 3788–3796 (2006)

Some of the massive blocks making up the great pyramids of Giza in Egypt (pictured below) are not limestone, but a synthetic mix like concrete, argue materials scientists.

The paper by Michel Barsoum of Drexel University in Philadelphia, Pennsylvania, and his colleagues is the latest entry in a decades-long argument. Most Egyptologists reject the idea, put forth in the mid-1980s, that the pyramids contain concrete.



A. VAN ZANDBERGEN/LPI

Barsoum's team took a fresh look at 15 samples using scanning- and transmission-electron microscopes. The samples contain ratios of elements, such as calcium and magnesium, that do not exist in nearby limestone. The imaging also revealed regions of amorphous structure. Both observations suggest that other substances were added to make a concrete mix, say the authors.

BIOFUELS

Positively negative

Science **314**, 1598–1600 (2006)

A mixture of prairie grasses could produce a biofuel that mops up more carbon than it produces, say David Tilman and his colleagues at the University of Minnesota in St Paul.

The grasses are 'carbon negative' because each hectare of the crop can lock away up to four and a half tonnes of carbon dioxide in soil and roots every year. This dwarfs the 0.3 tonnes or so of carbon dioxide released during production and combustion of the fuel.

Tilman concludes that the grasses would reduce carbon dioxide 6 to 16 times more effectively than grain ethanol and biodiesel, two rival biofuels.

JOURNAL CLUB

Thomas Mrsic-Flogel
Max Planck Institute of
Neurobiology, Martinsried,
Germany.

A neuroscientist asks whether neurons are enough when it comes to learning.

One of my main scientific interests is understanding how the brain adapts to and learns from experience. By 'brain' I normally mean networks of neurons, because the electrical impulses they produce are the common currency of sensation and perception; learning involves changes in the synaptic connections between these cells driven by sensory experience. But it's beginning to seem that the brain's plasticity depends on more than neurons alone.

A few months ago, I found myself debating with a colleague what role glial cells might play. I was sceptical. Some years ago, the various sorts of non-neuronal cells that go by the name of glia were thought of simply as a glue (which is what glia means in greek) that holds the brain together. More recently, though, these cells have been shown to form extensive networks important for regulating the local brain environment and to communicate with neurons.

Despite knowing all of this, I had not considered glia as serious players in neuronal plasticity. My thinking changed after reading two recent studies showing that glia not only change their activity after sensory stimulation (X. Wang *et al. Nature Neurosci.* **9**, 816–823; 2006) but also influence the strength of synaptic connections on neurons via a secreted soluble protein (D. Stellwagen and R. Malenka, *Nature* **440**, 1054–1059; 2006). Because the amount of secreted factor depended on the level of surrounding neuronal activity, these results may provide a glial link between sensory experience and synaptic plasticity. The challenge is now to show this directly in the intact brain. I am willing to give it a try — the glia may have changed my mind!

NEWS

When good cholesterol turns bad

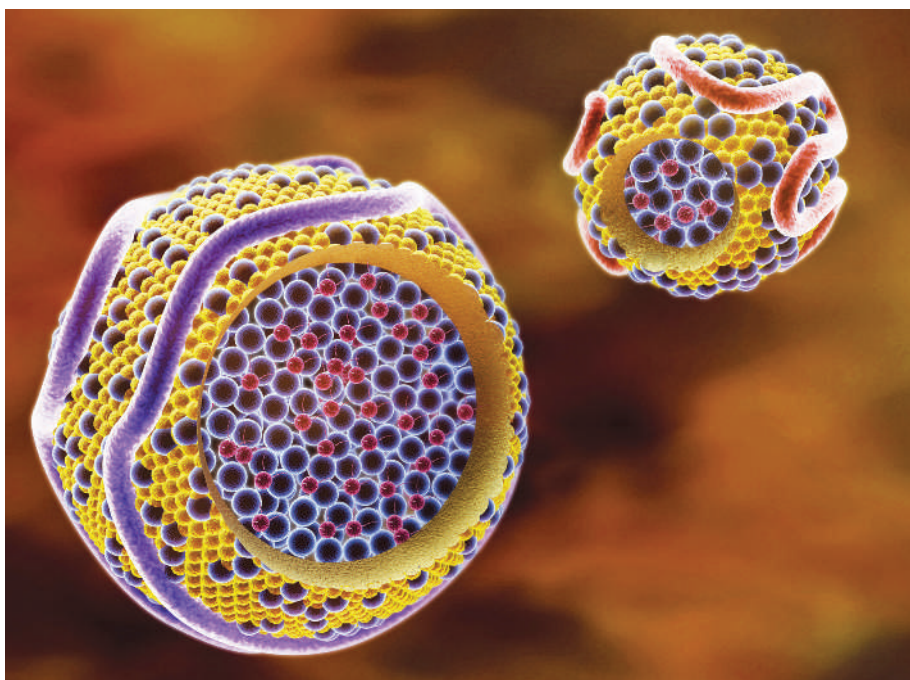
Last week's dramatic failure of a drug trial aimed at boosting 'good' cholesterol is bad news for Pfizer, the world's largest drug company. But will it also dash the hopes of those looking for a new way of preventing heart attacks?

Until now, the cholesterol story has been portrayed as a simple one, with just two characters: 'bad' cholesterol, which promotes heart disease, and 'good' cholesterol, which protects against it. Driving down levels of the former helps the heart, as should boosting levels of the latter.

The first idea is broadly accepted, thanks to statins, the best-selling drugs that cut levels of bad cholesterol (more properly called low-density lipoproteins, or LDLs). But with Pfizer being obliged on 2 December to abandon trials of its drug candidate torcetrapib, which was designed to increase levels of good cholesterol, the second idea has suffered a serious setback.

The trial in 15,000 people was called off because of the increased number of people who had died after taking the drug and a statin for 18–20 months, compared with a group taking the statin alone; 82 people died on the combined therapy, compared with 51 on the statin. "This cancellation came as a complete shock," says John Chapman, who has studied torcetrapib at the biomedical research institute INSERM in Paris.

But not everyone was so surprised. Some specialists have been harbouring doubts that



HYBRID MEDICAL ANIMATION/SPL

Some experts are questioning the benefits of raising the blood levels of 'good' cholesterol.

certain ways of raising good cholesterol — high-density lipoproteins, or HDLs — might cause more problems than they solve.

Bumping up HDL levels makes sense on the basis of many epidemiological studies showing that people with low levels of HDLs tend to be at higher risk of heart disease. But Pfizer's trial

was one of the first to test whether this beneficial effect could be mimicked using a drug to specifically boost levels of HDLs.

HDLs labour away in the bloodstream like a huge fleet of garbage trucks, carting off cholesterol to the liver for safe disposal. LDLs do the opposite, dumping their cholesterol load

Changes at the helm fuel Salk expansion plans

SAN DIEGO

The prestigious Salk Institute for Biological Studies is on the brink of change. Fresh leadership at the California research facility in La Jolla is set to direct it towards a major expansion.

The search for a president has already begun after neuroscientist Richard Murphy announced on 8 December that he will leave next July, after nearly seven years in the post.

And Irwin Jacobs, an electrical engineer who led his company, San Diego-based Qualcomm, to pre-eminence in mobile-phone technology, assumed the nine-year

chairmanship of the Salk board last month. Billionaire Jacobs is a popular choice. He has a stellar international business reputation, a track record in philanthropy, and is the first board chair to come from the Salk's home town.

"These developments create the opportunity for the Salk to unlock its substantial potential," says Kenneth Chien, a Harvard physician who has a Salk adjunct faculty position. "This is a very opportune moment to capitalize on a science-driven agenda."

Although it has only 62 faculty members, the Salk has set a sterling record for answering some of the

most basic questions in biology, and its strength and direction are closely watched.

The leadership changes come as the institute prepares to complete an expansion first envisaged in the 1960s, when polio researcher Jonas Salk founded the institute. Whereas three other non-profit research institutes in San Diego's cluster recently accepted lavish financial payouts to open labs in Florida, Salk officials spurned its suitors (see *Nature* 442, 729; 2006). "Three or four Florida groups asked us to consider moving," says Murphy. "We were not interested. Our strength is

remaining small and interactive."

The Salk plan calls for a new building for labs; a research facility for photonics, mass spectroscopy and human stem-cell projects; a child-care centre; and ocean-view residences for visiting scientists. Salk officials hope to win the necessary government approvals by next June, while launching a funding drive for the project. Costs are estimated to exceed \$100 million.

The institute is considered a national architectural landmark — it was designed by Louis Kahn and crowns an ocean-front cliff. So the expansion is already drawing



OXYGEN BURST SEEN BEFORE THE BIRTH OF COMPLEX LIFE

Air to breathe may have spurred evolution.

www.nature.com/news

GETTY

in blood vessels where it can accumulate and potentially clog the vessels. Torcetrapib works by inhibiting a protein called cholesteryl ester transfer protein (CETP), which moves cholesterol from HDLs to LDLs. This, it was hoped, would help clear more cholesterol from the bloodstream.

As Wall Street frets over the implications of Pfizer's failed trial (see page 805), cardiovascular specialists are equally abuzz about why the failure occurred. They want to know whether the problem lies with this drug alone, whether all drugs that inhibit CETP might be prone to similar problems, or whether meddling with HDL levels is inherently dangerous. The last possibility would have serious repercussions for researchers and drug companies already investing heavily in an array of other methods to boost or alter HDL levels.

The reason for the increased deaths in the trial participants will not become clear until the results of the abandoned trial are released. One hint lies in evidence from earlier studies that torcetrapib raises blood pressure in some people. That might have proved fatal in patients with already raised blood pressure. If this is the case — or if the problems are caused by some other factor unique to torcetrapib — then other drugs being developed to inhibit CETP might not suffer the same downfall. One drug being developed by Roche, for example, is thought to interact with a different part of the CETP molecule and might not affect blood pressure.

But some researchers are asking whether the very concept of inhibiting CETP is flawed. Normally, when the HDL trucks dump their cholesterol at the liver, they do so by offloading

some of it to LDLs on the way. But CETP partially prevents the transfer of cholesterol from HDLs to LDLs, so there has been some debate about whether drugs such as torcetrapib could cause HDLs to accumulate large amounts of cholesterol that they cannot traffic out of the body normally. "You're blocking a physiological mechanism to eliminate cholesterol," says Moti Kashyap, who directs atherosclerosis research at the VA Medical Center in Long Beach, California.

According to other experts, the idea of boosting good cholesterol may be far too simplistic to start with. HDLs are complicated particles, they point out, composed of different proteins and lipids. Each particle's components and chemical modifications are thought to affect how it behaves, and some recent studies have suggested that certain variants of HDLs may actually promote cardiovascular disease.

In one such study, a team led by Ben Ansell at the University of California, Los Angeles, studied people who defy medical expectation because they have particularly high HDL levels, but still develop heart disease. In lab assays, the team found that HDLs from these people behave in ways that could promote the build-up of cholesterol in the arteries rather than protect against it (B. J. Ansell *et al. Circulation* **108**, 2751–2756; 2003). Similarly, Ansell says, Pfizer's drug might have favoured the formation of damaging or dysfunctional forms of HDL. He adds that his finding is "very analogous to what I'm sure happened in this trial."

"One hint lies in evidence that torcetrapib raises blood pressure in some people."

Following this line of thinking, HDLs themselves should be reclassified into 'good' and 'bad' forms. This means that total HDL levels could be a poor measure of its actions and that heart drugs would work only if they boost the correct form of HDL. "It might not be enough to have a lot of HDL — what matters is how that HDL functions," says Prediman Shah, director of cardiology and atherosclerosis at the Cedars-Sinai Medical Center in Los Angeles.

Some researchers are optimistic about efforts to boost one of these particularly 'good' breeds, a variant of a protein found in HDL. The protein, called ApoA-1 Milano, was found in residents of a rural Italian village who have very low HDL levels but do not seem to be prone to cardiovascular problems. In 2003, a team led

by Steven Nissen of the Cleveland Clinic in Ohio showed that five infusions of this protein in humans shrank plaques in the arteries (S. E. Nissen *et al. J. Am. Med. Assoc.* **290**, 2292–2300; 2003). The approach is now being pursued by Pfizer.

Many cholesterol specialists say they are still optimistic that, by one means or another, tweaking HDLs can add to the benefits already accrued by statins. But at least some admit that the case for raising good cholesterol is nowhere near as watertight as that for lowering bad cholesterol. "There are many ways to raise HDLs. The problem is we really don't know if any of them work [to prevent the cardiovascular disease]," says Christie Ballantyne, a cardiovascular specialist at Baylor College of Medicine in Houston, Texas.

Helen Pearson

fire. A local design organization has criticized the plan, even though much of the new building will be underground, temporary structures and parking lots will disappear, and no final model has been decided.

Having Jacobs as board chairman should help considerably, though, especially given his services to the community — his many donations include \$110 million to the University of California, San Diego, and \$120 million to the San Diego Symphony.

"We plan to focus on raising money for the institute's scientific work," says Jacobs. The number of faculty to be added with the expansion has not been decided, but a 10% increase has been suggested.

Murphy's tenure had been a fruitful one. During his term, 16 faculty members were hired, major research centres were created, about a third of the labs were renovated and the endowment rose. "It's been a great six years," says Murphy. "I thought it best to go before the institute started the capital campaign."

Looking to the future, faculty head Inder Verma says: "For the first time, we have a board chairman who is from San Diego; this is very positive. He is a community leader with academic credentials." On the presidential search, Verma notes: "The faculty would prefer someone from outside, to bring in new blood."

Rex Dalton



An architectural landmark, the Salk Institute is set to grow.

P. ABRAHAMIAN/CORBIS

Agency set up to tackle bioterror

WASHINGTON DC

In a late-night flurry of activity as it rushed to adjourn, the US Congress last week created a new agency that will allocate \$1 billion to companies that are developing drugs and vaccines to tackle bioterror and pandemic agents. Blearily-eyed lawmakers also passed legislation governing the broad workings of the National Institutes of Health (NIH) and authorizing a hefty budget boost for the bio-

medical agency over the next two years.

Both bills — which President Bush is expected to sign into law as soon as this week — faced long odds as the Republican-dominated Congress entered its final hours before ceding control of both the House and the Senate to Democrats in January. But the determination of Republicans to leave on a note of achievement, and the insistence of Joe Barton (Republican, Texas), who chairs the energy

and commerce committee, that the NIH bill must be passed by the Senate for the popular bioterror measure to be brought to a vote in the House, combined to push both measures through. The complex horse-trading also involved a state-run health-insurance programme for children.

The biodefence bill establishes a new agency — the Biomedical Advanced Research and Development Authority (BARDA) — as part

SNAPSHOT

'So far, so good' for shuttle

NASA is breathing another sigh of relief after an apparently successful shuttle lift-off on the evening of 9 December. As *Nature* went to press, NASA officials were still checking Discovery's heat shield for any damage, but the night launch, the first since the 2003 Columbia disaster, seems to have passed without incident.

"So far, so good," was the assessment of flight director Tony Ceccacci. Small pieces of foam debris and ice fell off the shuttle's external fuel tank during the launch, but this was expected and they did not appear to strike the shuttle.

The night-time launch meant that NASA officials were unable to capture on-board video images with as much detail as they had on previous launches. The team is still investigating four 'low momentum' readings from the leading edges of the craft's wings, although they are confident that these are not evidence of debris strikes.

On 11 December, Discovery docked with the International Space Station, where it is delivering a new crew member, Sunita Williams, and an \$11-million extension to the station's solar-power system. Michael Hopkin



NASA

of the Department of Health and Human Services. The agency will initially get \$1 billion over two years to fund civilian research on countermeasures to bioterror and pandemic threats.

Senator Richard Burr (Republican, North Carolina), who sponsored the bill, describes BARDA as “an aggressive venture capitalist” that will partner universities and companies to help them through middle- and late-stage drug development and manufacture. The bill had broad bipartisan support, including co-sponsorship by Senator Edward Kennedy (Democrat, Massachusetts), whose state is a hub of biotechnology firms, and Senator Hillary Clinton (Democrat, New York), whose constituents suffered the most in the attacks of 11 September 2001.

The BARDA bill has long been a darling of industry, which had lobbied for its passage. Its backers noted that the \$5.6-billion Project BioShield, which was passed into law in 2004, provides funds for the government to buy vaccines and drugs once they are ready for use. At earlier stages in drug and vaccine development — stages the industry has termed the ‘Valley of Death’ — companies are therefore left to push forward at their own expense.

“This legislation recognizes that the ‘Valley of Death’ remains a barrier to effective countermeasure product development,” Jim Greenwood, a former Republican congressman who is president of the Washington-based Biotechnology Industry Organization, said in a statement after the Senate passed the measure to create BARDA.

Other proponents argue that developing agents to tackle what are only potential attacks and pandemics is far riskier than investing in a new diabetes medicine, for example, and that it’s reasonable to expect the government to share that risk. “It’s like a B2 bomber,” says Francesca Cook, vice-president of policy and government affairs at PharmAthene, a company based in Annapolis, Maryland, that is developing prophylactics and treatments against anthrax and chemical nerve agents. “Only the Department of Defense is going to buy that.”

But critics say the bill is a blatant giveaway to business, a security risk and a wasteful move in an era of ballooning federal deficits.

“I don’t know when this terrible cycle of splurging on biodefence is going to end,” says Ed Hammond, director of the US office of the Sunshine Project, a bioweapons watchdog group based in Austin, Texas, and Hamburg, Germany. “I don’t think we need nearly 20,000 people at over 400 institutions researching

biological weapons agents.”

Hammond refers to numbers compiled by the US Centers for Disease Control and Prevention in Atlanta, Georgia, which reports periodically on the number of people and institutions registered to handle potential bioweapons agents. He worries that the sheer number of people involved in such research increases US vulnerability to domestic terrorism, and encourages other countries to enter a race that could lead to the offensive use of biological weapons.

Fiscal conservatives also bemoaned the BARDA bill’s passage by a Congress that has spent \$26.7 billion on civilian biodefence since 2001. “The creation of a new office is always suspect,” says Tom Finnigan, a spokesman for Citizens Against Government Waste, a Washington-based group that monitors government spending, “because it will be duplicating roles that already exist. And it will never go away. Even if they stockpile all the vaccines they need, the office won’t close down. It will expand its mission, inevitably.”

The other bill passed in the small hours on Saturday was the National Institutes of Health Reform Act of 2006. For the first time since 1993, it offers broad guidance and new requirements for the structure and running of the NIH. Barton, the bill’s author, noted that even as the agency’s budget doubled between 1998 and 2003, its “hodge-podge” structure was left untouched.

The bill caps the number of institutes at the current 27, and requires the agency’s director to submit biennial reports to Congress cataloguing all NIH research and justifying the individual institutes’ priorities. It requires the NIH to set up a user-friendly electronic system to uniformly code all its grants and activities. And it establishes a ‘common fund’, comprising up to 5% of the total NIH budget in any given year, for research involving more than one NIH institute. But the most eye-catching provision is a funding boost that would increase the biomedical agency’s budget by more than \$4 billion to \$32.8 billion in 2008.

But there is no guarantee that the money will materialize. That remains in the hands of the congressional appropriating committees that fund the NIH year by year. “We recognize that it’s not money in the bank and that we’ll have to convince appropriators to do the same,” says Howard Garrison, a spokesman for the Federation of American Societies for Experimental Biology. “But this is a good sign. It’s a strong endorsement of the NIH.” ■

Meredith Wadman

ZOO NEWS

Locals in the English town of Borwick, Lancashire, were baffled when a seal turned up in a grassy roadside verge — more than 6 kilometres inland and some 1.5 kilometres from the nearest river. Despite its apparently heroic overland journey, the disorientated trekker, now nicknamed Sid, bore not a scratch.



RSPCA

ON THE RECORD

“Maybe we’ll find out that spaceflight turned out to be a historical aberration, like Zeppelins.”

John Pike, director of scientific monitoring group GlobalSecurity.org, on whether projects such as the International Space Station might lose impetus.

“Given... the risks of a pandemic originating from the Hajj, mandatory influenza vaccination for all pilgrims should be considered.”

An editorial in the *BMJ* ponders the flu risks of the annual gathering of some 2 million Muslims in Mecca.

NUMBER CRUNCH

25% is the cut in cod quotas to be imposed for 2007 by the European Commission after commissioners said that independent experts did not provide “clear quantitative scientific advice”.

0 is the unequivocal amount of cod the International Council for Exploration of the Sea (ICES) advises should be taken from European waters in 2007 to best preserve stocks.

5 is the number of years that the ICES has been advising that the quantity of cod caught in European waters should be zero, without being heeded by the European Commission.

Sources: The Guardian, BMJ, Environment News Service



EARTH SCIENCES
Read our blog from the
American Geophysical
Union meeting
<http://blogs.nature.com/news>

E. MAYNARD/ALAMY

Australia lifts ban on cloning

Australia has ended its ban on research that involves therapeutic cloning — giving researchers there a chance to join the small number of groups engaged in the procedure worldwide.

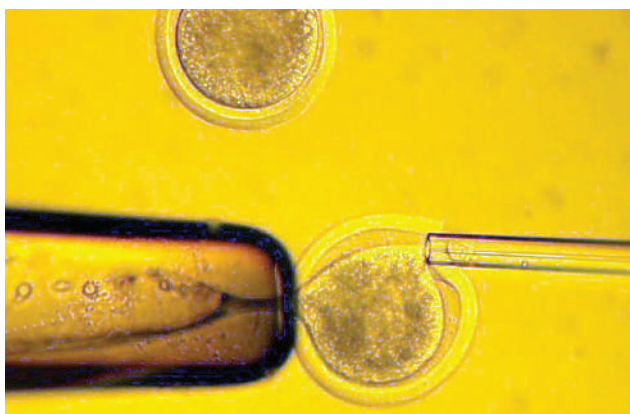
A bill lifting the ban was passed by the Senate last month, and by the lower house on 12 December — despite being opposed by both prime minister John Howard and the leader of the opposition, Kevin Rudd.

“I was losing confidence that we had the support,” says an elated Alan Trounson, director of the Monash Immunology and Stem Cell Laboratories in Melbourne. “But were able to win — even against the most powerful figures.”

Therapeutic cloning refers to the extraction of stem cells from embryos created by cloning, sometimes known as somatic-nuclear cell transfer. It can potentially create stem cells from patients with diseases such as Alzheimer's, multiple sclerosis or diabetes. Trounson hopes to use such stem cells to study the progression of these diseases and their possible treatment.

The bill repealed this month — the 2002 Prohibition of Human Cloning Act — banned the creation of cloned embryos for either reproduction or research. In 2004, Australia supported a US-led effort to have the United Nations ban all forms of cloning.

But now Australia becomes one of several countries, including Britain, Finland, Singapore and South Korea, that specifically permit the creation of cloned embryos for research



Nest egg: somatic-nuclear cell transfer gets the go-ahead down under.

purposes under certain conditions. The United States has no law stating what work researchers may or may not do — but they cannot use federal funding for therapeutic cloning. US researchers nonetheless account for about half of the groups known to be pursuing research on therapeutic cloning.

In 2005, a committee chaired by John Lockhart, a Sydney lawyer, reviewed Australia's 2002 cloning act. Its findings, which stressed the potential value of therapeutic cloning, are reflected in the new bill.

But the same bottleneck that has limited research elsewhere in the world — the low availability of human eggs — will also slow therapeutic cloning in Australia. Eggs are needed to reprogramme donor cells to an embryonic state. Women who donate eggs must undergo the same invasive procedure as those who have *in vitro* fertilization (IVF),

which carries a slight risk of serious side-effects (see *Nature* 442, 606–608; 2006). Australia will forbid direct payments to egg donors, and its researchers will have to rely on altruistic donations or on eggs from ovaries that have been removed for medical reasons. “There is not going to be an avalanche of research in the field,” says Tejia Peura, a stem-cell researcher at Sydney IVF, a fertility treatment company. She thinks that hers is one of three or four groups, at most, that will take up the research.

But the bill also opens up other avenues of related investigation. Parthenogenetic activation of eggs,

in which eggs start development without a sperm, will be allowed, for example. So will research involving the transfer of cytoplasm, the material surrounding the nucleus of an egg. The latter could lend insight into why eggs age, and, perhaps, offer hope for families with an abnormality in the cytoplasmic mitochondrial DNA that causes Leigh's disease in children. The bill will, however, prohibit the production of human embryonic stem cells by injecting human donor cells into animal eggs.

Details of how the change will be implemented, including licensing requirements for researchers, are expected early in the new year. Trounson says he cannot wait to get started, but adds that the widespread support the measure attracted has also been a huge boost for scientists' morale. “It was a great show of democracy in Australia,” he says.

David Cyranoski

Stem-cell technique ‘contrary to public order’

MUNICH

A German court has revoked a patent on a method for generating a class of human embryonic stem cells. The 5 December ruling is seen as yet another setback for stem-cell research in a nation where it is already constrained by tight regulation.

The federal patent court in Munich heard a charge brought by Greenpeace that the patent, held by University of Bonn neurobiologist Oliver Brüstle on a way to generate

precursor nerve cells, was ‘contrary to public order’. The environmental group argued that the derivation of the cell lines in question had involved the destruction of human embryos, which breaches guidelines issued by the German Patent Office. The judge, Eva-Maria Schermer, quickly ruled in Greenpeace's favour — much to the dismay of Brüstle, who arrived in court with three bodyguards to protect him.

The ruling will become binding in a few weeks. It was made in the

wake of a public call by Germany's largest research agency, the DFG, for a relaxation of stem-cell laws, which are stricter than those of many other European countries (see *Nature* 444, 253; 2006).

Brüstle says he will now appeal to the Supreme Court in Karlsruhe, arguing that the ruling goes beyond German law, which allows the use of human embryonic stem-cell lines created before 2002. Just last year, the ministry of research awarded Brüstle a grant for work with human

embryonic stem cells derived before 2002. He points out that the ministry stipulates that its grantees should attempt to patent inventions from projects that it supports.

But a ruling by the Supreme Court in Karlsruhe could take years. “In the meantime, it is very hard on me and my family,” Brüstle says, “particularly for kids whose father has been accused of doing things so bad they are considered to be against public order.”

Alison Abbott

Martian gullies turn tide in hunt for life

Planetary scientists are eager to step up their search for life on Mars after news that liquid water seems to have flowed down gullies (pictured left) in the planet's mid-latitudes within the past few years.

Mike Malin of Malin Space Science Systems in San Diego, California, and his team saw gullies forming on two crater walls on Mars between 1999 and 2005. The light colour of the gullies and their characteristic shape have convinced Malin that they were carved by flowing water (M. C. Malin *et al. Science* **314**, 1573–1577; 2006).

Gamma-ray observations show that plenty of ice is present on and beneath the surface of Mars. But with an average surface temperature of -63°C and an atmospheric pressure of less than 1% of Earth's, significant amounts of liquid water have been considered unlikely, to say the least. Malin is not suggesting that there are lush lakes, or waterfalls flowing down the red planet's cratered surfaces. More likely are brief spurts of rapidly evaporating water that push a mudflow down a slope.

Malin's peers are taking the news seriously. "These results are certainly impacting our science objectives," says Jennifer Heldmann, a geomorphologist at NASA's Ames Research Center in Moffett Field, California. In 2005, Heldmann suggested that gullies on Mars could be caused by water flowing under current martian conditions (J. L. Heldmann *et al. J. Geophys. Res.* **110**, EO5004; 2005).

Since Malin announced his results, Heldmann has arrived in the Utah desert, an area thought to have similar geomorphology to that on Mars. Heldmann will be visiting several sites in the desert over the next two weeks to work out more about the formation mechanism of gullies there, and relate this to the Mars results. "We're trying to tie together the gully observations with the formation mechanism," she says, "exactly what we are also trying to do on Mars."

HiRISE, the High-Resolution Space Imaging Experiment that is onboard the Mars Reconnaissance Orbiter (MRO), will also be following up Malin's results. Alfred McEwen, principal investigator for HiRISE, says that the instrument will take images of the gullies

identified in Malin's paper several times over the next few years.

It's very unlikely that flowing water will be seen in action. If there, it probably flows very rarely, and for only a few hours in small, isolated spots. Malin's back-of-an-envelope calculation suggests that HiRISE has a tiny 1-in-10²¹ chance of seeing water flow.

But HiRISE might be able to shed more light on exactly how the gullies are formed. McEwen accepts water flow as a reasonable hypothesis, but argues that other mechanisms, such as dust deposits, can't be ruled out. (Malin counters that dust slides appear as dark, not light, patches.) The MRO also has a spectrometer on board — the Compact Reconnaissance Imaging Spectrometer for Mars, or CRISM — which could detect the salts or frost that Malin believes are causing the light tone in the images.

But the only way to find out for sure is to land on Mars and go to the gullies. Landing on or near the bumpy craters, however, would require pinpoint expertise. Matt Golombek, landing-site scientist of the Mars exploration programme, says that none of the landers being considered at the moment has that capability. "It's not clear in the immediate frame how we're going to respond to [Malin's] discovery," he says.

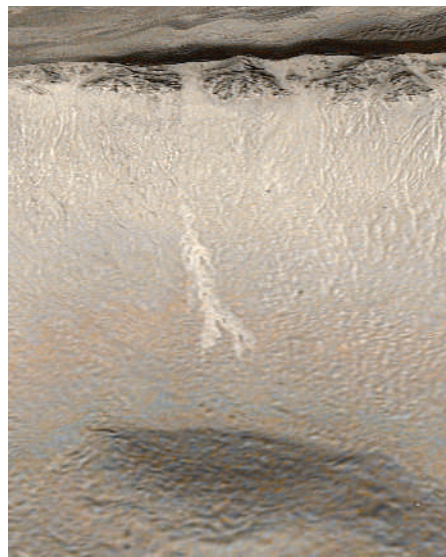
The European Space Agency is also unlikely to change its immediate plans. Its next Mars mission, called ExoMars and headed by Jorge Vago, is scheduled to launch in 2013. It aims to land in a flat area and drill into the rock to look for organic matter, among other things. Attempting to land near the gullies would escalate costs enormously.

But NASA has recently called for proposals for a scout mission to Mars. Malin has proposed a rover to land on plains near the crater, then drive down into the gullies. And avid Mars-watcher Phil Christensen from Arizona State University in Tempe has suggested a mission that would crash-land into ice-rich areas of the red planet's mid-latitudes and fly a craft overhead to monitor any debris.

Christensen is convinced that the gullies were formed by water. But he disagrees strongly about its source. Although Malin thinks that the water is coming from underground

NASA/JPL/UNIV. ARIZONA

NASA/JPL/MSSS



The appearance of light-coloured gullies on Mars suggests water has flowed there in recent years.

aquifers, Christensen says any underground water would have leaked out and evaporated millions of years ago.

That's a puzzle that Malin has yet to solve. He has several arguments for why some of the water below the surface might be liquid: salts might have lowered its melting point, or the subsurface of Mars may be warmer than previously thought. But it's unclear where the water is coming from. Malin speculates that it comes from melting surface ice or that it has existed in the planet's deep interior since Mars formed.

Christensen favours a model in which packets of surface snow melt when the climate warms. Other experts are divided. McEwen suspects both Christensen and Malin are right, whereas Vago is not so sure whether Mars has the right conditions to keep underground water liquid.

Either way, the finding that groundwater is seeping out of Mars is hugely significant for those who hope one day to find life on the planet. Liquid water is essential for life — at least for anything like life on Earth. And if life does exist on Mars, the only place it is likely to be able to survive is underground, and that is out of reach of our probes. If liquid water from aquifers is making it to the surface, it might carry microorganisms with it — although once there, they are unlikely to survive more than a few seconds. "It means that in the search for life, water is accessible and at the surface," says McEwen.

Malin points out that, so far, there is no evidence of biological activity on Mars. But he adds: "Our results certainly point to a place where biological activity might occur." ■

Katharine Sanderson

A vision of life after Blair

Tony Blair's choice of successor as Britain's prime minister has seemed little more than a formality since he announced in September that he would stand down. So UK scientists have begun to wonder what a government led by Gordon Brown, the current chancellor of the exchequer, would be like.

A few hints emerged last week, when Brown unveiled his annual pre-budget report. UK researchers have rarely questioned Brown's commitment to research, given that he has boosted science spending by 70% to £2.5 billion (US\$4.9 billion) since 1997. But his largesse comes with a desire to micro-manage and a close eye on what industry wants. So it is little surprise that Brown's report reinforces his commitment to science, while raising fears about how he might govern it.

For example, the report restated an intention, first aired in March, to ensure that the Department of Health's research funds actually get spent on research. At present, a sizeable fraction of the department's £600-million research budget gets diverted to the National Health Service.

But Brown's blueprint for this reform looks complex and perhaps slanted in industry's favour. Funds for the Medical Research Council and the health department would be coordinated by a new body, the Office for Strategic Coordination of Health Research. Another new organization, the Translational Medicine Funding Board, would work with these to get results from bench to bedside, raising the possibility that basic-research funds could be raided to pay for work that industry wants done.

Reforms to another mainstay of UK science have also received a mixed response. Every seven or so years, UK researchers undergo the Research Assessment Exercise (RAE), an expert review of university research used to distribute more than £1 billion of government money to higher-education institutes. Pleas for reform of this time-consuming activity have grown louder in recent years, prompting Brown to launch a consultation on a metrics-based alternative earlier this year. With the results now announced, some critics may be wishing they had stayed silent.

Measures of external research income, postgraduate training and research impact, the last by citation analysis, are broadly

in line with what researchers wanted.

The plan also fits the treasury's desire to streamline the process. But the proposals would almost completely eliminate peer review. Bibliometrics experts say properly weighted citation statistics produce the same results as expert review, but many, including the Royal Society, do not agree.

Ole Petersen, a cell biologist at the University of Liverpool who chaired the society's RAE panel, uses the story of Nobel prizewinner Bert Sakmann to explain why. Sakmann's key paper on ion channels was published in 1976, but citations didn't take off until new experimental tools came into use a decade later. If metrics had been used to evaluate Sakmann's paper in the mid-1970s, they would have returned a low score. "But everyone knew this was a breakthrough," says Petersen.

Even Brown's oft-repeated statement that science and the innovation that springs from it should be central to Britain's economy could prove controversial. Some suggest he may create a Ministry for Science to push the idea. But this could mean shifting more funds from basic to applied research, as Brown did with £60 million of university money last week.

Brown's spending increases have turned Britain's labs around, but some fear that the future may look less certain if applied research is seen as the *raison d'être* of science. "You always have to be wary when people march down that road," says Peter Cotgreave, director of the Campaign for Science and Engineering, a London-based lobby group. "They can forget about the science base."

Jim Giles



How will UK science fare under Gordon Brown?

T. MELVILLE/REUTERS

US public unperturbed by skills shortage claims

Researchers may be worried about the United States' scientific performance, but the public is not, a recent poll suggests.

A survey of 1,000 registered voters by the American Council on Education, a Washington-based higher-education group, found that 85% think the most significant threat to US economic competitiveness is

cheap labour abroad, not the development of a more skilled workforce overseas. The poll also found that only 31% believe that maths and science classes are "very relevant" to non-science majors, and only 54% believe that students should take more maths and science courses.

Scientific competitiveness has become a rallying cry in Congress and the White House, both of which supported this year's American Competitiveness Initiative, boosting funding to education and some

fields of research (see *Nature* 439, 644–645; 2006). But the latest poll shows a disconnect between political rhetoric and public opinion, according to David Ward, the council's president.

Survey dispels health threat from mobile phones

In one of the biggest and longest-running studies yet of mobile-phone use and cancer, researchers have found no evidence for a link between the two.

Unlike earlier studies, the researchers relied on records from mobile-phone companies rather than owners' unreliable memories to gauge phone usage. They tracked subjects for an average of 8.5 years, with some followed for up to 21 years. Joachim Schüz of the Institute of Cancer Epidemiology, Copenhagen, and his colleagues looked at more than 420,000 mobile-phone users and checked them against records at the Danish Cancer Registry.

Electromagnetic radiation from the phones' antennas can penetrate a little way into the brain, prompting fears that they might cause cancer, but the researchers found no increased incidence (J. Schüz *et al.* *J. Natl Cancer Inst.* 98, 1707–1713; 2006).

Solar cells break the record for energy efficiency



A California company has made the world's most efficient solar cell. Engineers at Spectrolab, based in Sylmar, built the cell by growing three different layers of gallium arsenide crystals on a wafer (pictured). Each layer contains a different dopant, which allows them to absorb different parts of the Sun's spectrum. When coupled with a device to concentrate sunlight, the cell can convert 40.7% of the light into electricity. Most commercial solar cells are around 15% efficient.

The new cell broke the previous record of 39.3% efficiency. Although a 1.4% increase may seem small, says Spectrolab president David Lillington, "every extra per cent of efficiency helps you generate more power over the lifetime of the system".



Andrew von Eschenbach: permanently in charge at the US Food and Drug Administration.

Senate vote ends months of waiting for FDA head

The US Senate last week finally confirmed Andrew von Eschenbach as commissioner of the Food and Drug Administration (FDA).

Although von Eschenbach was nominated in March, the vote to confirm his position had been blocked by two Republican senators, Chuck Grassley (Iowa) and David Vitter (Louisiana). Senate leaders used a procedural vote to override the pair.

Von Eschenbach, a urologic surgeon, is a former director of the National Cancer Institute and has been acting head of the FDA since September 2005.

The confirmation means that the \$1.9-billion agency once again has a permanent head — a situation it has enjoyed for just 18 months during George W. Bush's tenure in the White House.

NIH researcher confesses to breach of ethics

An Alzheimer's researcher at the US National Institutes of Health (NIH) pleaded guilty on 8 December to a criminal conflict-of-interest charge.

Trey Sunderland, who led a geriatric unit at the National Institute of Mental Health, has been a focal point of congressional inquiries into allegations of conflicts of interest at the NIH.

Federal prosecutors said that Sunderland accepted \$285,000 in consulting fees from the drug company Pfizer without approval from his NIH bosses. They added that Sunderland failed to properly report travel expenses and consulting income from Pfizer that "directly related" to his federally funded research.

Having struck a deal with prosecutors, Sunderland faces two years of probation, 400 hours of community service and a fine to be set by the judge. He will also have to repay the government the \$300,000 in fees and travel expenses he collected from Pfizer.

President quits California's stem-cell institute

California's \$3-billion stem-cell research institute is in need of a new head following the resignation of president and chief scientific officer, Zach Hall.

Hall, who has led the California Institute for Regenerative Medicine (CIRM) since its creation by state ballot in 2004, announced last week that he intends to retire some time in the next six months. "My job was to get things started," Hall says.

His departure comes as the CIRM is, indeed, making some progress as a grant-making agency after being mired in litigation (see *Nature* 437, 800–801; 2005). A state loan of \$150 million and nonprofit-financed bonds have allowed it to begin issuing research grants even though legal action against it is continuing.

Correction

The Special Report "Anti-evolutionists raise their profile in Europe" (*Nature* 444, 406–407; 2006) stated incorrectly that the University of Leeds plans to introduce remedial courses on evolution for first-year science students. The courses are in fact part of a core lecture module on evolution, introduced to give second-year students sufficient evidence and confidence to debate evolution in schools, universities or society in general.

BUSINESS

Aerial ambition

A spin-off from a UK university has established a solid niche as a leading supplier of microsatellites. But taking the company to the next level is a challenge. **Geoff Brumfiel** reports.

The phrase 'mission control centre' may conjure up images of banks of NASA engineers pondering giant display screens in Houston, Texas. But at Surrey Satellite Technology, southwest of London, mission control is a modest affair. From here, a sparsely furnished room in a nondescript business park, a dozen or so desktop computers are quietly keeping tabs on several working spacecraft. "When TV stations come visit us, they like us to put on headsets and pretend to look busy," says Phil Davies, an engineer who manages the company's American business. "They've all seen too many James Bond movies."

The empty control room sums up the unassuming way that Surrey Satellites has conducted its business, since it was founded in 1985 by Martin Sweeting, an aerospace engineer at the University of Surrey in Guildford. Since then, the company has built a business making microsatellites — spacecraft weighing 10–100 kg — for customers who want to go into space but don't want to spend more than US\$10 million for the privilege.

Surrey Satellites' revenues have grown considerably since 2000 (see graph). "We have identified a market of some \$1 billion over the next five years for small satellites," says Sweeting, now the company's chief executive, "and we plan to grow our share." By 2016, he thinks, the company could achieve annual sales of up to £100 million.

Satellite craft

But steeper competition, regulatory hurdles and a shortage of cheap launch vehicles suitable for the company's brand of hardware all stand in the way of Surrey's aspirations.

In Europe, larger contractors are dipping their toes in the microsatellite market. "These firms have woken up to the threat of small satellites eating some of their lunch," says Sweeting. And in the United States, the company's opportunities are constrained by a law requiring that publicly financed satellites go into orbit on expensive US rockets. Customers have been uninterested in cheap satellites when their launch is going to cost several hundred million dollars, company officials say.

Surrey Satellites' initial clients were mainly national governments with little experience in space: it built satellites for South Korea, Portugal, Chile and others for uses ranging from telecommunications and military surveillance to crop monitoring and disaster management. The company also trains engineers from client firms in satellite construction and operations.

The privately held company is one of the most successful of about a dozen microsatellite companies around the world, with 24 launches to date. Other players include AeroAstro, based in Ashburn, Virginia, and SpaceDev in Poway, California. "What sets Surrey apart is efficiency

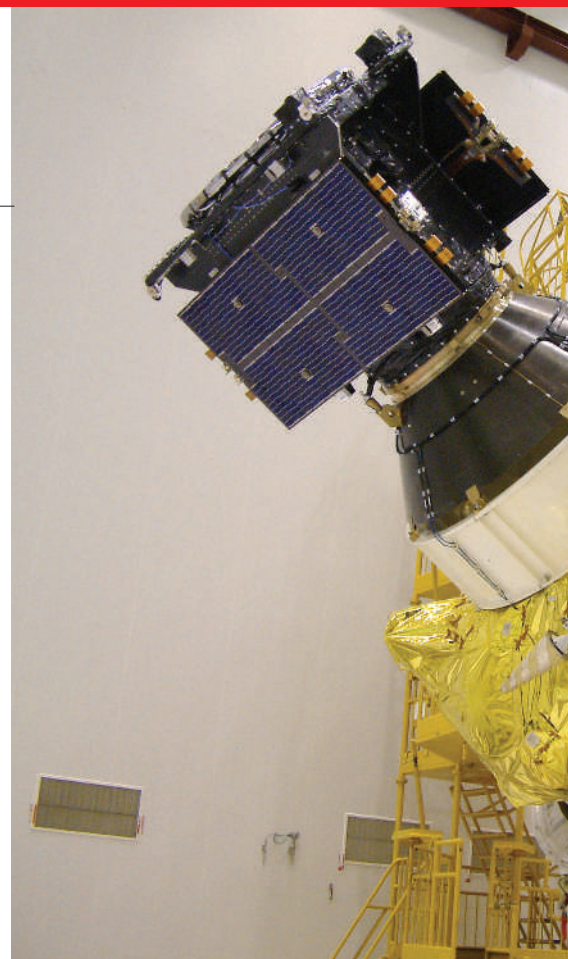
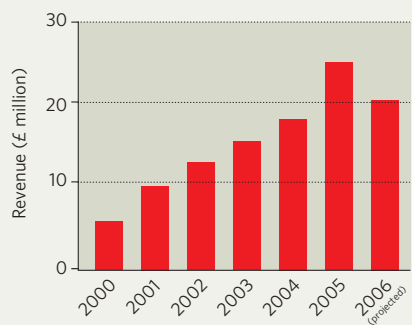
and pragmatism," says Jonathan McDowell, an astronomer and space-industry watcher at the Center for Astrophysics at Harvard University in Cambridge, Massachusetts. The company uses off-the-shelf camera components, computer

disk-drives and processors, along with standardized designs to build new satellites to clients' specifications in as little as 12 months. They're simple, reliable and look almost indistinguishable from each other — but can carry out fairly sophisticated tasks nonetheless.

In the past few years, Surrey Satellites has moved from its small-nation niche to try to establish itself with larger clients in Europe and the United States. In 2003, for example, it was selected to build a test satellite for the European global positioning system, Galileo. The product, GIOVE-A, was a behemoth by Surrey Satellites' standards, weighing more than 600 kg. The company has also built satellites for Alca-

"We have identified a market of \$1 billion over the next five years for small satellites."
— Martin Sweeting

SURREY SATELLITES' ANNUAL SALES



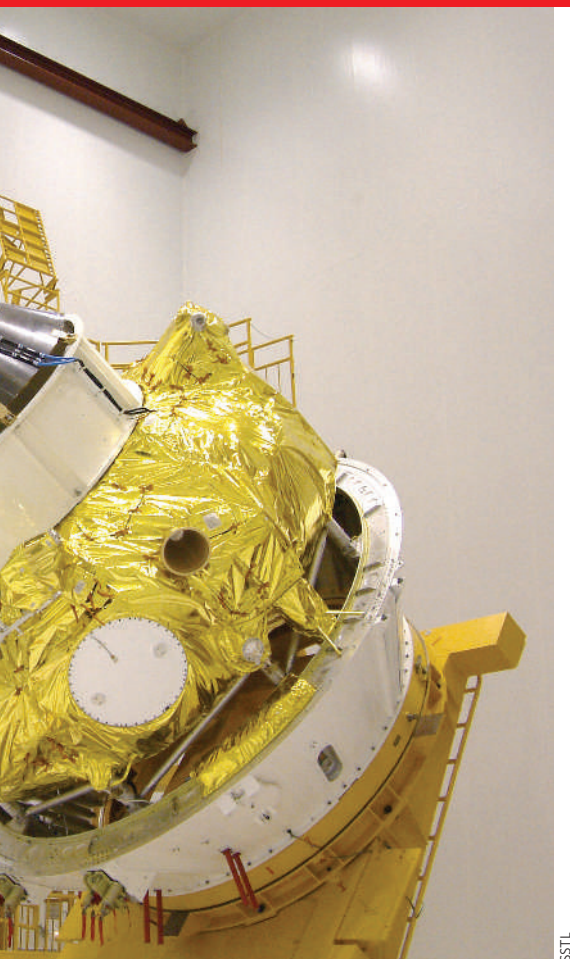
tel, the French telecommunications company, and for the US Air Force. "As the capability of our satellites has improved, the customers have got more and more serious," says Davies.

The transition to larger markets has put Surrey Satellites under strain. The University of Surrey, which holds 80% of what is now a substantial, high-risk company, is trying to divest. So it has sent the company on the hunt for new, private investors. Last year, Elon Musk, founder of California-based Space-X, bought a 10% stake: Sweeting says more investors could come on board towards the end of next year.

Stellar competition

The shift in emphasis is also bringing the company up against more-established rivals with deeper pockets. Last year, for example, it lost a European Space Agency (ESA) contract for a constellation of small, Earth observation satellites to Astrium, the space division of EADS, the Franco-German parent company of Airbus. The ease with which the larger firm outbid Surrey Satellites perturbs Davies: "We were perplexed at how Astrium can provide three satellites for €86 million (US\$114 million), when that would normally be their price for one," he says. Astrium spokesman Jeremy Close denies that the company underbid for the contract, claiming that ESA give the contract primarily on the basis of technical capability. It had "nothing to do with price," he says.

Competition is set to heat up further as the US federal government gets serious about



SSTL

Sky's the limit: the GIOVE-A navigation satellite was a step up for Surrey Satellites.

microsatellites, says Carissa Christensen, a space analyst with the Tauri Group, a consultancy based near Washington DC. The Department of Defense, for example, has said that it wants a capability to quickly launch customized satellites to monitor specific regions of interest — or to spy on other satellites. If branches of the Pentagon decide to invest big in small satellites, she says, this could “create a much more robust set of competitors than Surrey currently has”.

So far, the company has struggled to establish a presence in the United States, mainly because of the 1988 Act. Additionally, stringent US export control rules make it tricky for satellite customers in the US government to talk directly to overseas contractors about their technical requirements.

So Surrey Satellites is working to build partnerships in the United States that will give it better access to the world's largest satellite market. This April, it signed a marketing agreement with the US branch of defence contractor BAE Systems. Davies hopes it will bring the contacts and legal knowledge needed to nurture better relationships with US customers. And the Surrey Satellites' newest investor, Space-X, hopes to develop cheap launch vehicles for possible use with its satellites in the United States. Taken together, company officials hope, these moves could crack the US market. ■

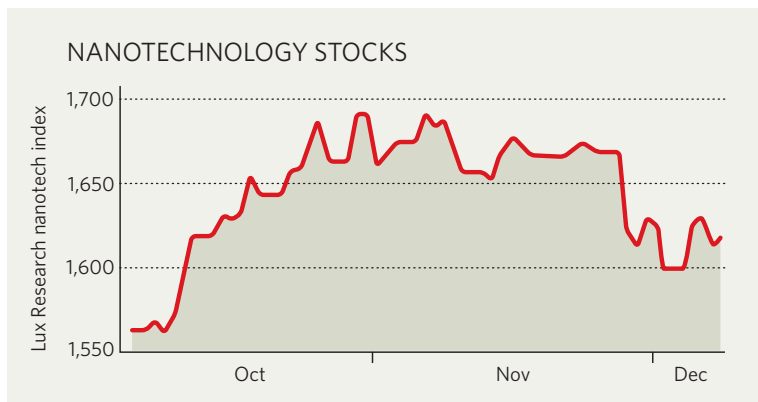
IN BRIEF

Pfizer setback The world's largest drug company is facing some turbulence after a large human trial of its cholesterol drug candidate torcetrapib was abandoned. The 2 December announcement that the trial would end, due to an unexpectedly high death rate among its participants, caused Pfizer's shares to fall from almost \$28 to \$25, knocking about \$20 billion off its market capitalization. Analysts say that the New York-based company may be forced to revisit its entire corporate strategy as vital revenue earners — such as cholesterol drug Lipitor, the world's biggest selling drug, which torcetrapib was intended to replace — come off patent.

Swiss role Oerlikon, a Zurich-based engineering company, has won a SFr320 million (US\$268 million) to equip a huge solar-cell factory in Offenbach, Germany. The plant, which is being built by Saudi-backed solar-cell producer API, will be able to build enough cells to install 160 megawatts of electricity-generating capacity each year. The order reflects growing demand for solar cells: Oerlikon said in a statement that its solar business — which uses cheaper, amorphous silicon instead of the pure crystals used on most photovoltaic cells — was “exceeding all expectations”.

Not ticking Shares in British biotechnology company Evlutec plunged by three-quarters after its drug candidate rEV131, which is extracted from tick saliva, failed in a human trial for the treatment of hay fever. The company, a spin-off of the Centre for Ecology and Hydrology in Oxford, UK, is continuing to test the drug for other uses, including the treatment of eye inflammation after cataract surgery (see *Nature* 439, 533; 2006). Mark Carnegie Brown, the company's chief executive, said the trial failure was “tremendously disappointing”.

MARKET WATCH



Nanotechnology stocks in the United States have held up well in the past two months — unless, that is, they had the misfortune to be associated with silver nanoparticles.

Shares in Nucrust Pharmaceuticals — a company based in Wakefield, Massachusetts, that uses such particles in its popular wound dressings — took another hammering after the US Environmental Protection Agency (EPA) said it would regulate washing machines that used silver nanoparticles to kill germs (see *Nature* 444, 520; 2006). Nucrust uses such particles in unrelated applications. But its shares — already sharply down after a failed September dermatology trial — took a tumble late last month and are now trading at a quarter of their summer high.

The Lux Research index, which contains a basket of stocks producing

and using nanotechnology, has nonetheless held steady. Peter Hebert, chief executive of New York-based Lux Research, which compiles the index, says that despite a weakening US economy, corporate investment in nanotechnology continues to expand. He adds that the EPA ruling has been misinterpreted and won't hurt nanotech as a whole.

The end of the year saw better showings by Flamel Technologies of Lyon, France, whose shares surged after the US Food and Drug Administration approved a GlaxoSmithKline heart drug delivered using Flamel's microparticle system, and have almost doubled to \$32 since the summer. And shares in Oregon-based FEI, which makes scanning electron microscopes, rose sharply in November on good quarterly results.

Colin Macilwain



S. CLARKE/REX FEATURES

A MATTER OF LIFE AND DEATH

Few issues in biology have the power to fire up emotions like the topic of animal research. The public face of the debate is largely characterized by polarized positions: the use of animals in the laboratory is simply either 'all good' or 'all bad'. The imagery and arguments tend to be quite crude. On the one hand members of the public are confronted with images of angry animal-rights protesters picketing labs and offices, calling for an end to all animal experimentation. On the other, they are presented with claims from supporters of animal research that the use of such creatures in the lab is essential and should not be questioned.

Somehow, the voice of the middle ground has been lost. With stories appearing in the news of scientists and their families being threatened by extremists, it is not surprising that many researchers are reluctant to talk publicly about their work. But, as Emma Marris reveals on page 808, those who are prepared to discuss the issue add a range of highly nuanced views to the debate. To explore this further, *Nature* conducted an anonymous survey of 1,682 biomedical scientists, and found a simi-

larly broad range of opinions. Interestingly, the survey also suggested that many scientists feel that they receive insufficient support from their universities or institutions to allow them to speak on the subject more openly (see page 789 and the full poll results online at www.nature.com/news/specials/animalresearch).

Bench scientists are not the only ones who can find themselves in the firing line. On page 811, Kerri Smith talks to a vet, whose job it is to care for experimental animals, about the challenge of balancing the pressures from both anti-vivisectionists and researchers, while trying to improve the lives of lab primates. On page 812, David Cyranoski speaks to researchers working on primates who are worried that, in addition to violent animal-rights protests, regulations are an increasing threat to their work. And on page 814, Jane Qiu looks to the future of mouse research and how the advent of large-scale genetics projects will affect the use of this rodent. ■

To voice your opinions in this debate, visit *Nature's* newsblog at www.nature.com/news.

GREY MATTERS

Many scientists have nuanced views on animal research. But they are rarely heard, says **Emma Marris**.

Many readers of *The Guardian*, a British newspaper, will have been surprised at a recent online article in which Sophie Petit-Zeman, a neuroscientist and journalist, explained how she could at the same time be a vivisectionist and a vegetarian¹. But they will not all have been surprised in the same way. Some will have been surprised at the existence of such a complex position in an area where much of the discussion is depressingly black and white. Others will have been surprised not by the position itself, but by its being discussed in public.

As a poll of *Nature* readers working in the biomedical sciences reveals, many scientists who work on animals have complex takes on the issue. But they are not often willing, or encouraged, to express these feelings. Some of this is directly due to fear of animal-rights extremists; some is an indirect effect of the polarized atmosphere that surrounds the issue. In some labs, at least, scientists feel pressured to keep quiet about the grey areas of debate, lest they undermine the official mantra

And how do they square the ethics of it all? As well as polling our biomedical readers (for full results see <http://www.nature.com/news/specials/animalresearch>), *Nature* set out to get some voices from the front lines — and found a lot of ducked heads. It would be fair to say that the average researcher prefers not to talk to the press about his or her work. And yet, there are those who are not only willing to talk, but have plenty to say. It quickly becomes clear that each researcher has his or her own system of ethical equations in place, but that the simplified pro-con debate makes it very difficult to communicate this — or have any kind of calm conversation about animal research.

So let's meet the researchers. Tom Burbacher runs an infant primate lab at the University of Washington in Seattle, which models the cognitive effects of prenatal exposures to environmental contaminants in macaques. He talks publicly about his work, but does not defend the entire enterprise of animal research, believing it is a waste of time to defend such a large, abstract concept. He adds that he feels that his work, which is clearly linked to human health, is easier to explain than some blue-sky research, such as mapping the brain or describing how vision works. "There is a lot of basic research going on that is harder to talk about," he says.

Burbacher got involved in toxicology as an undergraduate, and has been working with

animals ever since. "I had a study that followed animals from the time they were born until they were more than 20 years old. I got old with them," he says. Then, they were killed. "It was tough," he says. Burbacher is not shy about the emotional toll that working with animals can sometimes take. He says the toll is two-fold: "It does wear on you sometimes. I have on different occasions thought about getting out. The death of an animal is an acute stress, but the activists are a chronic stress."

Fine lines

Chris Harvey-Clark, director of the Animal Care Centre at the University of British Columbia in Vancouver, has even greater day-to-day contact with animals, working with research subjects from sea lions and hummingbirds to

transgenic mice. Harvey-Clark, like many in similar positions, was once a private-practice veterinarian, and remains a confirmed animal-lover, saying he often feels emotionally closer to his charges at the centre than to his old clients' pets. In the face of their inevitable deaths, Harvey-Clark must work to retain his humanity and empathy. "How do you keep caring for the animals without being scorched by the fact that you are using them up?" he asks.

His answer lies in the fact that suffering and death have long characterized the relationship between humankind and the animal kingdom — from animal predators that prey on humans, to the slaughter of wild and farmed animals by humans for food. As Harvey-Clark puts it, "without a farm background, it is hard to understand that you can both care for things



and also understand that they are going to wind up being food, or data”.

It's a perspective shared by many scientists, including Cynthia Otto, a clinical veterinarian and researcher, who is also a disaster-scene veterinarian with the Federal Emergency Management Agency. She has cared for search-and-rescue dogs in the immediate aftermath the terrorist attacks of 11 September 2001 in Manhattan, and has gone looking for pets left behind by Hurricane Katrina in New Orleans. She kills rabbits in research, but she also treats pet rabbits as a vet. “I also eat meat. I respect what the animals can give me. If I was not going to do any animal research, I would not eat meat or wear leather, but in a certain way, I feel like that would be less respectful — not taking advantage of what they can give us.”

But somehow, the complexities of arguments such as these seem to have been lost. “Whenever you talk about the research, the stock answer is to say that we are curing cancer or saving premature babies. You don't talk about finding out what a bit of the brain does just because you are

quite curious to know,” says one UK researcher, who works with rats, but prefers not to be identified because he worries that he might be targeted by animal-rights activists. “I have heard animal-research advocates say that you have to say everything as nicely as possible, and that edges towards fabrication. They say that everything heads towards a cure for something and that all the experiments work. We should be more honest because a lot of people will still go along with it.”

Finding a voice

The same researcher feels that some scientists go so far as to imply an anguish they don't really feel about the practice. “I think the standard thing that people tell you is that they are some sort of tortured soul that can't sleep at night but have to do it to cure cancer. I do feel bad about using rats but I can sleep at night — we use the fewest possible and try our best to ensure they suffer as little as possible. People have been pushed into portraying themselves like that because of animal-rights activities.”

The silence — or spin — from the researchers may be most profound in the United Kingdom, where radical animal-rights activists have perhaps the longest and most disruptive history. Colin Blakemore, a neuroscientist who has long been the target of animal-rights extremists' activities and who now heads the UK Medical Research Council in London, says that this strategy is flawed. “We are perceived as being distant and unwilling to speak and therefore as having something to hide,” he says. Blakemore has encouraged universities to let their faculty members talk freely about their work and supported more public scrutiny of animal research.

But Blakemore doesn't necessarily believe that there is much hidden discussion or soul-searching about animal research inside the field, apart from complaints about the bureaucratic acrobatics involved in working with animals. “It is not the kind of thing that's always talked about in the coffee room,” he says. And yet he has his own shades of grey. Although he vigorously defends the practice, he is relieved to be done with his years working directly with animals, despite his consistently vigorous defence of the practice. “The longer I used animals, the less comfortable I was with it,” he says. It's a sentiment he believes is common in the research community. “I don't know of a

single scientist who would not prefer to use alternatives if they were available.”

Burbacher believes that talking about the details and benefits of his work is the best policy. Activists have left black roses on his doorstep and mobbed his office, but he keeps talking.

“The more people

“It is hard to understand that you can care for things that are going to wind up being food, or data.” —

Chris Harvey-Clarke

know about what is going on in the lab the better,” he says. But even he is circumspect when he first meets people. “If I don't know the people very well, I usually tell people I teach. You can usually tell whether the animal part should be brought up or not.”

Marin Stephens, vice-president for animal-research issues as the Humane Society in Washington DC, believes that the polarized debate has harmed communication between researchers and mainstream groups such as his organization, which favours a reduction in animal research. “There is not a lot of cross-talk,” he says. “There is a lot of stereotyping and demonizing on both sides.” He adds that the Humane Society co-founded a confidential group of stakeholders from both sides that is



D. ALLISON

D. ALLISON



meeting to discuss the issues and to try to undo some of this polarization.

"In my own experience, I have found a diversity of opinion in the research community," says Stephens. "Some care greatly for animals and are very helpful in changing the status quo. Others see things in black and white and have no use for people like me." And researchers aren't the only ones being unfairly lumped into one position. Stephens says that the Humane Society has sometimes been grouped with disruptive extremists for political reasons. "If you pin down individual researchers, they would admit that there are different perspectives in the different groups," he says, "but what one hears quoted is that they are all of a piece."

To find out more, *Nature* conducted an anonymous survey of 1,682 readers working in the biomedical sciences. What we found supported the claims quoted here: that there is a diversity of opinion within the community about animal research, with many respondents calling for more dialogue among scientists and with the public. The majority of respondents, just over 70%, believed that the animal-rights movement had made voicing a nuanced view of animal research in public more difficult for individual researchers. Some commented that it also hindered debate within the science community of how to best do animal research.

Otto says she has experienced this dampener on discussion first-hand. She is trying to develop alternatives to using animal models in her speciality, the study of sepsis. She wants to move away from disease-model animals to real sick animals in veterinary practice. She believes they will be better models, but part of her motivation is to help individual animals. "Naturally occurring diseases might be a better reflection of what is going on in a human than inducing something in a mouse or a rat.

People's pets that spontaneously develop these overwhelming infections need the treatment, and we would be more responsible because we are not actually creating diseases in animals."

Otto says she's had a hard time getting a hearing for her proposal from researchers resistant to change. And she's not alone. Ian Roberts, epidemiologist at the London School of Hygiene and Tropical Medicine, put forward the idea in a recent article in the *British Medical Journal* that all animal-research studies should be preceded by systematic reviews of their clinical utility, as human trials commonly are². To illustrate the helpfulness of doing so, he and his co-authors surveyed some reviews that said many animal studies were of dubious worth.

Although there were some considered objections to the paper, including one from Blakemore³, there were also some angry responses from both sides. "One of the things that made the argument more difficult," Roberts says, "is that in animal experiments, views are so polarized that you can't get any serious discussion about the methodology. When we first published, we were accused of being anti-vivisectionists, and then of being vivisectionists. The whole debate seems to cause people to become emotional and irrational."

Rational responses

So what can be done? Discussion groups such as the one Stephens attends are one potential solution, as is the Boyd Group, an independent UK forum founded in 1992 to produce solid information about animal research and bring those of differing opinion together. The group was founded by Blakemore and Les Ward, an animal-rights activist. Ward has since left the

group, which he says has become stalemated. But he believes the group was useful in that it was one of the few places where moderate activists and moderate scientists sat down and talked things over.

"I want to see the total end of animal experimentation, but I am not stupid enough to think that it is going to happen overnight," says Ward. "Everyone has to be willing to move their position." Ward says that scientists rarely sought him out to talk about their qualms about research, but when he visited labs, he was swarmed. "It was clear to me that they wanted to speak," he says. "They were intimidated, but claiming intimidation is also a dodge to avoid having to speak."

Another inclusive effort was the Nuffield report⁴, a two-year effort to capture the state of the debate run by the UK-based independent Nuffield Council of Bioethics. Barry Keverne, chair of the Royal Society's animal-research group, says it was one of the best. "They produced a huge document — it has all members of society in it, lawyers, historians, philosophers, scientists and anti-vivisectionists. Most scientists have no problems discussing this openly with people who have a genuine interest."

In Europe, the moderate approach is more established. Vera Rogiers is a toxicologist at Vrije University in Brussels, Belgium, and the chair of ECOPA, the European Consensus-Platform for Alternatives, which promotes the development of alternatives to animals in research. She says that in the rest of Europe, the discussion is more "healthy". "Most people are very well aware of the dialogue and the development of alternatives. We ask that the people who are around the table sign a statement that

they accept the three Rs — reduce, replace and refine. Usually all the people that we have around the table have no problem with that."

The old adage may truly fit here: there are as many views about animal research as there are thoughtful people. But as long as the debate is played out as a ping-pong game between hard-core activists and hard-core defenders, anyone in the middle who stands up to be heard risks getting hit.

Emma Marris is a reporter for *Nature* based in Washington DC.

"I don't know of a single scientist who would not prefer to use alternatives if they were available."
— Colin Blakemore

1. Available online at: http://commentisfree.guardian.co.uk/sophie_pettizeman/2006/08/confessions_of_a_vegetarian_vi.html.
2. Pound, P. et al. *Br. Med. J.* **328**, 514–517 (2004).
3. Blakemore, C. & Peatfield, T. *Br. Med. J.* **328**, 1017–1018 (2004).
4. Ethics of Research involving animals by Nuffield Council of Bioethics (May 2005). Available online at: http://www.nuffieldbioethics.org/go/ourwork/animalresearch/publication_178.html.

CAUGHT IN THE MIDDLE

Researchers aren't the only ones who concern themselves with animal welfare in the lab. Vets are asked regularly to monitor and care for these animals — a role that can call for some difficult decisions. **Kerri Smith** talks to Sarah Wolfensohn, head of veterinary services at the University of Oxford, UK, about the challenges and conflicts presented by caring for experimental animals.

By law, UK institutions doing animal research must have a vet to oversee such work. What does your job involve?

The primary role is to look after the health and welfare of the animals. We have to do routine visits, and quality control of the animals' environments. We do a lot of health monitoring and advise licence holders on their project licences or experiments, refining them to reduce the impact on the animals. As part of that we are constantly advising on, for example, different anaesthetics or methods of collecting blood samples.

We also run all the training courses for licence holders and do research projects, trying to find objective measures of animal welfare. We are not in the position of accepting or vetoing programmes of work — my team feeds into the ethical review process of how welfare can be improved, but our role is an advisory one.

You specialize in the welfare of primates — which in Oxford generally means macaques. How have you been improving their lot?

The most significant difference was when my group got agreement from Oxford's research groups to include foraging in the primates' environment. It's now official — the National Centre for the Replacement, Refinement and Reduction of Animals in Research has just published guidelines on primate accommodation and care, which state 'all primates should be given the opportunity to forage daily'.

A primate in the wild will spend 75% of its time foraging for food — it's not natural to put it in a cage and give it a lump of food at meal times. Putting in wood shavings in which the food is hidden means they can forage around and look for it. It takes time, so they spend their time much more naturally, and they can use exploratory and social behaviours. It makes a huge difference to primate lifestyle.

This followed on from a general shift in practice towards housing primates together in



Animal well-being: welfare for macaques considers environment enrichment using cages outdoors.

groups. Twenty years ago, you might have had eight single cages in a room, each housing one monkey. Now you have all eight monkeys having the whole room. They can indulge in social behaviours and interact, and they can use the room three-dimensionally.

How do you train scientists?

In a number of ways. For example, we've made a DVD about how you can improve primate welfare by taking them out of cages and putting them into open rooms. It has made a significant difference, especially in the United Kingdom, where the majority of animals are held in groups.

You must come up against some fairly negative opinions.

It's always a challenge. As a vet surgeon you take an oath when you qualify that says that your constant endeavour will be for the welfare of animals committed to your care, but when you're dealing with experimental animals, their welfare has the potential to be compromised. That inevitably puts you in a difficult position. We have to balance the welfare against the quality of the science, and there are occasions when that is challenging. The scientific benefit may be

absolutely clear and outstanding or it may be rather more vague.

Are there other sources of conflict?

I went to a conference last year where somebody — not a lab animal vet — came to speak from the Royal College of Veterinary Surgeons, and he said: 'When I came here I was wondering why anybody would work in this field, and now I've spent a day talking to you all, I still don't understand why anyone would work in this field.' It's because you are caught in the middle all the time. The anti-vivisectionists don't like you, because you're on the other side, as they perceive it. Some scientists perceive you as trying to change the way they do their work.

The whole ethical review process and the legislation is seen as a hurdle by many scientists. What you do have to protect against is the minority of over-zealous scientists who are single-mindedly pursuing their scientific goal. Those are the problem people — those few affect everybody else.

What impact does being squeezed from both sides have on your work?

One of the problems with the anti-vivisection movement is that because it targets everybody involved, it's really difficult to recruit people to this area. The net effect is that the anti-vivisection movement is bad for animal welfare.

You constantly have to think, 'because I'm making a difference and I'm improving animal welfare, I have to ignore the anti-vivisection pressure, and think about the medical benefits that come out of this.' In order to have medical benefits, there will have to be some animal use, and therefore you have to do your best for the welfare of the animals that are being used, and that's why I do it. I know that over the past 20 years I have made a difference to the welfare of animals that have been used, despite the pressures from the anti-vivisectionists on one side and the scientists on the other.

Kerri Smith is a science writer currently in Nature's London office.

BFC FARM, ISRAEL

Primates in the frame

Primate researchers have long faced violent protests over their work. But in some countries, regulatory obstacles are taking a greater toll. **David Cyranoski** meets European scientists who feel that bureaucratic pressures are closing their labs.



Kevan Martin's monkey research project has been running for ten years. This year it hit a snag. Martin studies the brains of macaques to find out why higher-processing areas — the neocortex that accounts for 80% of the human brain — have been such an evolutionary success story. "If we can arrive at a general solution of the structure and function of the neocortex, it has enormous consequences," he says. The National Centre of Competence in Research, a Swiss government initiative, has reviewed and approved his work, supplying him with generous funding.

Martin's project at the Swiss Federal Institute of Technology Zurich has also been reviewed and approved every two or three years by the local animal experimentation committee. But this year the licence was delayed, forcing him to halt work in June. The eleven-member committee voted 5–4 against the licence, but the local chief veterinary surgeon overruled the decision, and the licence was granted in November. However, six members of the committee have now exercised their right under Swiss law to challenge the licence. The legal challenge has put Martin's work on hold indefinitely.

Klaus Peter Rippe is a philosopher at the University of Zurich who heads the local animal experimentation committee and is one of those protesting against Martin's licence. He attributes the extra attention on Martin's work, in part, to a change in the law. In December 2005, Switzerland reformed a law on animal welfare to protect the "dignity of creation" of animals. The vague wording has real effect,

says Rippe: "The legal framework changed; objections that were in the moral sphere are now in the legal sphere."

Martin thinks that the battle for his studies to resume will take at least another six months. "It's a relentlessly slow, drip-by-drip method of making life uncomfortable," he says. Martin has already seen one lab disappear, and fears that his will be next. Indeed, the number of primates used in research in Switzerland has dropped from 536 in 2000 to 408 in 2005.

The marmoset studies of Zurich colleague Joram Feldon, head of the behavioural biology laboratory, ended in April after questions were raised by the local committee. Feldon's study involved isolating marmosets from their mothers for 30–120 minutes repeatedly during the first four weeks of their lives to induce stress — a model for human depression that, Feldon says, cannot be matched by non-primate models.

The potential success of that model was what worried Rippe, whose local committee requested the ethical review of Feldon's work. "What would be the consequences if the experiment went well?" The worrisome answer: drug companies would start using more marmosets for research. So Rippe sought the advice of two national ethics committees, which formed a joint working group to evaluate Feldon's project. After weighing up the stress induced and the potential benefits to humans, the committees concluded that the severity level of the marmoset studies was ethically unacceptable.

Although the report does not have any legal jurisdiction over Feldon's licence, the

lead scientist on the project quit, and Feldon decided to focus on the experiments with rats and mice that make up the majority of his research. Martin, however, is not yet ready to give up: "I am the only person in Europe doing this research. If I'm gone, it's over," he says.

Martin and Feldon both point out that their lab's treatment of monkeys has not changed since their studies began some time ago. They found the unscientific nature of the ethics review particularly galling.

The working group comprised a biologist, an animal-welfare activist, a veterinary surgeon and three philosophers. In response to criticism that it did not include a primatologist, Rippe argued that they consulted four primatologists, including a member of Feldon's lab who was in charge of the marmosets. He also says that the 24-page report, which included lengthy discussions of ethical positions by Kant and other philosophers and contained no scientific citations, "is an ethical evaluation of primate research. The formal rules for ethical arguments differ from papers of the natural sciences."

Behind the scenes

Bombs, threats and arson perpetrated by animal-rights extremists grab the headlines, but behind-the-scenes manoeuvring may be taking a greater toll on primate research. Many researchers in European countries feel that they are being picked off one-by-one through regulatory mechanisms. Indeed, such indirect tactics may be more effective than direct campaigns.

Klaus-Peter Hoffmann, a cognitive neuroscientist who also works on vision, says that by forcing longer reviews and approval processes, or by getting them denied outright, opponents

"Objections that were in the moral sphere are now in the legal sphere."
— Klaus Peter Rippe



ILLUSTRATION: D. ALLISON

of animal research are succeeding in restricting primate research to a few centres such as the one in Tübingen. “No new laboratories are opening up in Germany,” he says.

Hoffmann, who will retire in two years, has previously gone to court to protect his right to do research. But, he says, there is little motivation to replace researchers like himself. “I don’t think the faculty is keen on primate research. The opposition is not only from the outside. Scientists also ask whether it is really necessary,” Martin, too, sees primate research slowly being strangled by regulations: “If you’re 35 years old and you have to stop for 18 months for approval, why would you bother?”

Primate research has always been the most controversial arena for the battle between animal-rights activists and researchers, and the need for it is growing. Ageing populations prompt more neurological studies into Alzheimer’s and Parkinson’s treatments, drug and vaccine developers say they require larger primate samples to improve preclinical trials, and stem-cell studies will need primates to bring the much-hyped research to the clinic. In 2002, some 52,000 primates were used for research in the United States and another 10,000 in the European Union.

Worrying trends

Researchers in Europe worry that their experiments will need to move overseas if, as they suspect, retiring researchers are not replaced, those coming from other countries are blocked and young researchers are discouraged from entering the field.

Alexander Thiele, a neuroscientist investigating vision at Newcastle University, UK, received a five-year €1.4-million (US\$1.8 million) fellowship from the Volkswagen Foundation to set up a laboratory in his native Germany. “You’d expect the university to be happy — they get a professor almost for free,” Thiele says. But his plans to move to the Charité Medical School at Humboldt University in Berlin were dashed in July, when its ethics review committee voted 4–2 against accepting him.

Thiele attributes the vote partly to university politics and partly to pressure from animal-rights activists. Before the vote, critical stories, based on leaked confidential information, ran in the local papers, says Thiele. Other universities have also turned Thiele down; he’s not too hopeful that he will return to Germany.

Meanwhile, in the United Kingdom, where protests against primate research have the longest, and most violent, history, researchers have seen a swelling of public support and legislative actions in response to the more extreme activities of animal campaigners. Simon Festing, director of the Research Defence Society, a group in favour of animal research, says that some of the more unsavoury actions of the animal activists are generating negative headlines for them. For example, activists dug up the remains of a member of a family that ran a guinea-pig farm, forcing it to close in August 2005. The public outrage was mostly directed against the activists.

In May this year, UK Prime Minister Tony Blair signed a petition in support of animal

research, which has to date been signed by more than 21,000 people. In a public statement, Blair noted that new powers for the police and courts would be used to counter the threat from animal-rights extremists.

Animal activists in the United Kingdom harass not only researchers, but also janitors, security guards or third parties who do business with animal researchers. They claimed victory, for example, in 2004, when the University of Cambridge scrapped plans for a large primate research facility. A BBC television documentary last month reported on a similar battle over the construction of an animal research laboratory at the University of Oxford. The documentary featured Oxford neurosurgeon Tipu Aziz, who experiments on Parkinson’s disease in monkeys and performs related surgery on humans. He says that he felt the need to speak out when construction at Oxford was halted last year. “It is my strong belief that the Cambridge site was not built because scientists didn’t educate people about what they’re doing,” says Aziz.

A fighting chance

But not all primate researchers are as determined as Aziz. This summer, Dario Ringach, a neuroscientist at the University of California-Los Angeles, announced that he would stop doing research on monkeys after the Animal Liberation Front (ALF) tried, unsuccessfully, to detonate a bomb on his colleague’s doorstep. The ALF victory may prove costly in the long run. In its wake, support for a pending Animal Enterprise Terrorism Act grew. The act, which will make it easier for law enforcement to crack down on economic harassment and impose greater penalties, was signed by US President George W. Bush at the end of November.

Frankie Trull, president of the National Association for Biomedical Research, a non-profit organization that supports the use of animals in research, says that the past four or five years have seen an increase in extremist activity. “I hear many stories of researchers having difficulty in finding postdocs,” she says. However, the new law “was a big victory for biomedical research”.

The battle is set to continue. Neither side looks ready to give up over the Oxford University facility, which is again under construction. This week, an independent UK review of the scientific case for primate research announced its support for such work, but suggested more can be done to improve animal welfare. Meanwhile, as Feldon watches his colleagues come under fire, he regrets that he did not fight to keep his experiment running. “Giving up might have been a mistake,” he says. ■

David Cyranoski is *Nature*’s Asian correspondent.

“It’s a relentlessly slow, drip-by-drip method of making life uncomfortable.”
— Kevan Martin

Mighty mouse

Coordination and integration of the results of animal research are an ever-increasing challenge. **Jane Qiu** finds out what happens when big biology meets a small rodent.

If mice could build a high-rise supercity, it might look something like this. Stack upon stack of clean, shiny plastic cages gleam inside the air-conditioned suites at the Wellcome Trust Sanger Institute in Cambridge, UK. With a capacity to house thousands of mouse lines, the institute has the ambition of becoming the world's largest warehouse for knockout mice, in which individual genes have been selectively deleted, or 'knocked out', from the genome. There are about 25,000 mouse genes, so the scale of the project is enormous and the cost is astronomical. "No single institute will be able to accomplish this alone," acknowledges Allan Bradley, director of the institute.

The Sanger Institute is part of an international effort to accomplish this feat, which promises to change the way mouse genetics is done for ever. "Biology is not a cottage industry anymore. We have to move away from thinking about one gene and one project at a time," says Steve Brown, director of the Medical Research Council's Mammalian Genetics Unit in Harwell, Oxford. "To face the enormous challenge of understanding the genome, we should be thinking about projects that will allow us to access every single gene."

It sounds ambitious, but such projects are very much a sign of the times. To crack the big questions about how something as complex as a mammalian genome works, biologists are increasingly turning to multi-million dollar projects that aim to capitalize on all the information coming out of genome-sequencing projects. Although researchers are certain that the mouse projects will yield a wealth of valuable new knowledge, questions remain about whether this approach is really the best one. "Although there is no doubt useful models will result, would the outcome of these knockout projects justify the cost and colossal number of animals involved?" asks David Threadgill, an expert on mouse models of human disease at the University of North Carolina at Chapel Hill. "That's the million-dollar question."

Certainly, such large-scale approaches can pay off. Last week, the Allen Institute for Brain

Science in Seattle, Washington, published the results of its three-year endeavour¹ that looked at the activity in the mouse brain of almost every gene in the genome — at a cost of US\$40 million. "This is a relatively small cost for the huge amount of information we have gained," says Ed Lein, a director of the neuroscience division at the Allen Institute. The resulting atlas provides significant insight into the brain's function and lets scientists measure the differences in the level of a gene's activity in different parts of the brain.

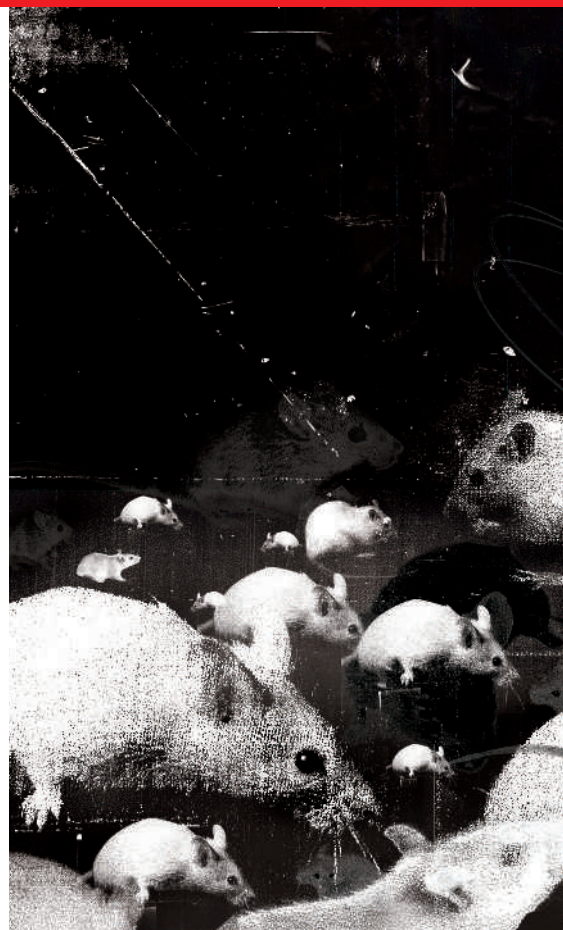
Big ambitions

The mouse mutagenesis schemes, however, are the most ambitious projects yet. Two research initiatives — the European Conditional Knockout Mouse Mutagenesis (EUCOMM) and the Canada-based North American Conditional Knockout Mouse Mutagenesis (NORCOMM) programmes — were launched in October 2005. They were joined a few months ago by the Knockout Mouse Project (KOMP), which was spearheaded by the US National

Institutes of Health (NIH). At a total cost of more than US\$70 million, the immediate goal of these projects is to turn the first steps to making knockout mice into a high-throughput enterprise by disabling genes in embryonic stem cells. Japan and China have also indicated an interest in developing similar projects. The teams plan to establish centralized archiving

systems to distribute reagents and mice to research communities around the world. This is only a prelude to the ultimate goal of generating and characterizing the entire cohort of 25,000 knockout mouse lines, which would cost at least another US\$600 million.

A further project, the European Mouse Disease Clinic (EUMODIC) will be launched next February with a handsome sum of €12 million (US\$16 million) for creating another 300 knockout mice and analysing 600 mouse lines in total, including those that will be generated by EUCOMM. Meanwhile, the Mouse Genetics Programme at the Sanger Institute is planning to create and characterize 250–500



knockout mice per year in the following decade. "Ultimately, we would like to have comprehensive genotype and phenotype databases for every single knockout," says Brown, who heads EUMODIC.

Geneticists are no strangers to attacking genomes with high-throughput, production-line approaches — they have been doing it for years with yeast and fruitflies, and even vertebrates, such as the zebrafish². But trying to pull off a similar trick in a mammal will be much more costly. Mutants of yeast and fruitflies are easy to produce and cheap to maintain. The generation of mouse knockouts is more involved and expensive, yet these costs are dwarfed by the price tag of maintaining the animals and characterizing their physical attributes, or phenotype.

Another big issue concerns the implications for animal welfare. Depending on individual genes and the method used to produce them, the number of mice needed to establish a line stretches from 50 to several hundred. On top of this, another couple of hundred animals are needed for basic analyses of genetic make-up and phenotype. So, for the 25,000 genes in the mouse genome, more than 7 million animals would be needed to generate and characterize all the knockout lines.

And the consequences of knockout mutations vary greatly, from small effects that are barely noticeable to debilitating conditions. "To closely monitor potential and actual suffering and distress in those animals on a large scale is very challenging," says David Mor-

"We hope that the knockout projects will illuminate gene function in the context of human disease."

— Colin Fletcher



PHOTOS: E. MATHIS/X. YU/D. THREADGILL; ILLUSTRATION: D. ALLISON

ton, a biomedical ethicist at the University of Birmingham, UK. In Europe, bodies that regulate animal research are committed to the principle of the Three Rs — refinement (to minimize suffering and distress), reduction (to minimize the number of animals used) and replacement (to avoid the use of living animals). Is the spirit of the knockout projects in line with this principle? “That is the issue researchers, funding agencies and regulatory bodies have to consider very carefully,” says Morton.

A balancing act

Proponents would argue that the benefits in terms of making large advances in biomedical knowledge balance these costs. And many, such as Karen Steel, director of the Mouse Genetics Programme at the Sanger Institute, say that even though a huge number of mice is needed, the coordination and technological efficiency of these projects mean that, in the long run, fewer mice would be used than if individual labs were left to their own devices. “This kind of high-throughput project will, in fact, reduce the number of animals used in research,” she says. “It will be a lot more cost-effective than what can be done by individual laboratories.”

The idea of an international mouse knockout project sprang from the difficulties that researchers had in generating or obtaining mutant mouse lines. Not all laboratories have the capacity to create knockout mice. And finding out which mutant lines are available is

not always easy, let alone getting hold of them. A recent survey conducted by the NIH³ shows that researchers have knocked out 11,000 mouse genes, most of which are not reported either because they belong to the private sector or because they do not have obvious phenotypes. Of the 4,000 unique knockout mouse lines that have been published, only 1,000 are in public repositories. As a result, there is a large duplication of effort: more than 700 knockouts are generated three times or more; in an extreme case, a mouse line is replicated 11 times.

“There isn’t a common mechanism in place to enforce deposition of knockout mouse lines into public repositories,” says Wolfgang Wurst, coordinator of EUCOMM and director of the Institute of Developmental Genetics at the German National Research Centre for Environment and Health (GSF) in Munich. Many knockouts were not generated in a standardized way or properly characterized, he adds. So, rather than trying to ‘recover’ mouse lines scattered around various laboratories, the researchers have come to the conclusion that the way forward is to generate a comprehensive and public resource of mutant embryonic stem cells — almost from scratch — by capitalizing on efficiencies of scale and a centralized production effort.

The enormous value of the mouse in biomedical research is not in dispute. Mice provide useful models for human disease and have helped researchers to make important breakthroughs in basic research in fields rang-

ing from molecular biology to behavioural science. “Nobody would work with animals unless it were the only way to address biological questions,” says Steel. “It’s much easier and cheaper to work with cells in a Petri dish, but that doesn’t necessarily tell you how genes behave in living organisms.” Instead, much of the debate about the mutagenesis projects centres around their cost-effectiveness and the validity of their scientific approaches.

To many researchers, it is one thing to generate a useful resource for a wider scientific community, and yet another to create and characterize 25,000 knockout mouse lines as the starting point to understand gene function. Some researchers doubt whether this is the most efficient way to obtain useful information about how the genome works.

The complexities of function

A key issue is that the function of genes is context-dependent — it depends on the genetic make-up of individual animals. “So it is not a matter of which genes do what, but how they behave as part of a complex genetic network,” says Allan Balmain, a cancer biologist at the University of California, San Francisco. Researchers often come across the problem that inactivating the same gene can yield very different phenotypes depending on the breed or strain of mouse. The strain used in research is therefore often crucial for the specific biological questions being addressed. So much so, in fact, that the Jackson Laboratory in Bar Harbor, Maine, launched the Mouse Phenome

Project in 2000 to collect baseline physiological and behavioural information for 40 different inbred strains of mouse, so that researchers could choose the most appropriate strain for their area of interest.

But it turns out that, for technical reasons, the different groups of the knockout project have used different strains of mice as their starting points. EUCOMM and NorCOMM are using the 129 mouse — the strain in which knockout technology was first developed. But this strain has some conditions that mean it is not usually chosen for a number of disciplines, including immunology and behaviour. Most researchers work with a strain called C57BL/6, the strain that KOMP has chosen to make its knockouts, although the technology to do this is less efficient than that for the 129 mouse. Once the technology is better established, EUCOMM and NorCOMM will switch to the C57BL/6 strain. It is not clear how easy it will be to compare results from the different strains.

A cog in the works

Another issue taxing mouse geneticists is whether these production-line systems will be sensitive enough to detect every aspect of a phenotype. In many cases, a knockout mouse does not show any obvious phenotypes, and researchers often have trouble knowing where to look for subtle phenotypes even when they have some idea what the gene does. This happens because genes often have overlapping functions, and can compensate for each other if one is lost.

To get around this problem, researchers frequently need to breed several knockout strains to generate double- and triple-knockouts, in which two or three genes are disabled simultaneously. “Using the knockout approach to study genes with unknown function on such a large scale may not be the most effective strategy,” says Threadgill. William Richardson, a neurobiologist at the Wolfson Institute, University College London, agrees: “Phenotyping is not a straightforward business. We have moved a long way from the big dream that knockouts will explain everything.”

But not everybody agrees. Martin Hrabé de Angelis, director of the German Mouse Clinic at the GSF, argues that knocking out a gene because you think it might be involved in a particular process can mean that new leads are missed. “Many researchers got their working hypothesis wrong and therefore focused on the wrong pathways,” he says. “And few would do



D. THREADGILL/UNC

Mouse supercity: warehouses of knockout mice could provide vital data about genes and their function.

comprehensive phenotype assays to study areas outside their specialities.” In his clinic, established in 2001 and a partner of EUMODIC, 240 parameters in 14 different areas are measured routinely as part of a high-throughput phenotype screen of mutant mice. Of the 50 knockout mouse lines studied, 42% of them have new phenotypes not detected before.

Some researchers are not convinced. “Phenotype studies are more than just getting a bunch of numbers,” says Richardson. Indeed, to uncover the mechanism for a particular characteristic, one needs to ask the right question. “It is a concern that we are going to end up with a lot of superficial phenotypes across many domains, but none of them is deep enough to answer difficult biological questions,” cautions

Threadgill.

At the end of the day, few would question the power of the knockout approach in deciphering gene function in the context of human disease. “The genome projects have generated a wealth of information in terms of gene sequence, but we have no clue what half of those genes do,” says Colin Fletcher, programme director at KOMP. “We hope that the knockout projects will illuminate their function in the context of human disease.”

But whether knockout mice can accurately mimic the complexities of human disease remains to be seen. The International HapMap Project has revealed more than 10 million polymorphisms — different versions of genes — in humans. And the Mouse Resequencing Project, which completed sequencing the genomes of 15 strains of mouse, has shown similar levels of complexity. These results are key to understanding individual variation in disease susceptibility and the effect of gene-environment interactions. “The main determinant of human genetic diversity is not which genes are or are not knocked out, but which genetic variants one carries in the genome. That is where the real interest of functionality lies,” says Balmain.

Indeed, many researchers are trying to build such variation into the animal models of human disease and to design screening strategies by targeting specific genetic variants. To those who have a taste for big biology, genome-wide association studies focusing on variations in genes may well be the next big thing. ■

Jane Qiu is a science writer based in London.

1. Lein, E. S. *et al.* *Nature* doi:10.1038/nature05453 (2006).
2. Driever, W. *et al.* *Development* **123**, 37–46 (1996).
3. Grimm, D. *Science* **312**, 1862–1866 (2006).

Nature has a commercial collaboration with the Allen Institute for Brain Science in the Neuroscience Gateway.

“Biology is not a cottage industry. We have to move away from thinking about one gene or project at a time.”
— Steve Brown

Cultural differences reduce Japanese researchers' visibility on the Web

SIR — As scientists, we are keenly aware that the world is developing into a single 'laboratory without walls', in which information passes as easily to the other side of the world as to the person working in a neighbouring institute. Although some people may be uneasy with this, to the brightest minds it is an enormous opportunity for progress, particularly in fundamental research. Yet information-sharing is not necessarily symmetrical, and depends on the tools that each contributor has available.

Our experience in managing the international research projects sponsored by the Human Frontier Science Program (HFSP) illustrates the problem.

This organization — of which we are president (M. I.) and secretary general (T. W.) — was established more than 15 years ago by Japan as an international programme for research into fundamental life processes. The HFSP secretariat, which is based in Strasbourg, selects postdoctoral fellows and research projects via international review committees.

Publication databases and powerful search engines allow the secretariat to update information regularly and to find it readily. However, it has become obvious that not all institutes and countries are on an equal footing in this respect (see www.hfsp.org/pubs/HFSP_articles/websites-scol.php).

In the case of Japan, it has become apparent that many scientists suffer from a lack of international visibility, in that they are very difficult to find by search engines and indeed in publication databases.

As a consequence, Japanese scientists are less likely to be invited to participate in collaborative projects or to become reviewers, which deprives them of a full international experience. Three main issues need to be addressed.

First, internationally comprehensible web pages must be constructed, to make a scientist's research interests, research group and publications immediately clear to anyone who visits the site. Many traditional Japanese-language scientists' websites start with a description of their philosophy and artistic interests, which in Japan are recognized as important in a potential mentor. Although this is culturally appropriate for Japanese students and postdocs, its relevance is, unfortunately, lost on the international visitor, who is accustomed to the succinct presentations typical of Western research institutes and universities. One simple remedy would be for Japanese researchers to have a Western-style page within their website, easily accessible and clearly

signposted, in English, on the homepage.

Second, many academic institutions have websites based on their curricula, which are appropriate for Japanese students, but are of limited interest for international visitors. It is important that the homepage of such institutional websites provides a clear option headed 'research', in the English language, that leads to a page summarizing the research in a style that is familiar to international visitors.

Third, in many regions of the world, numerous scientists have similar or identical family names and initials, making literature searches in PubMed very difficult or impossible. This is certainly an issue in some Asian countries, including Japan. Some concerted effort is necessary to resolve this problem — perhaps by the addition of laboratory codes, or a 'zip code' for the initials of individual scientists — to allow these scientists to compete fairly on the international level.

All of these are pressing issues in global science communication. Frontier-level international research is becoming concentrated in those institutions and laboratories that have the maximum visibility on the World Wide Web.

Masao Ito, Torsten Wiesel

International Human Frontier Science Program Organization, 12, quai Saint-Jean, B. P. 10034, 67080 Strasbourg-cedex, France

Nature welcomes comments from readers at Nautilus, our author blog: http://blogs.nature.com/nautilus/2006/12/web_visibility.html — Editor, Nature

Real-space solution to the problem of full disclosure

SIR — As discussed in Editorials in *Nature* and *Nature Structural Biology*^{1,2}, authors submitting research papers that describe molecular structures to Nature journals are required upon request to provide structural coordinates for reviewers to assess the quality of the work. But in the competitive field of structural biology, full disclosure of coordinates to anonymous peer review does raise serious concerns.

Nonetheless, the local nature of structure model quality — about which commonly provided global indicators such as R and R_{free} reveal no detail³ — in fact requires reviewers to be able to access and assess the quality in relevant areas of the structure (in particular around sites with bound ligands), which is not possible without disclosure of coordinates and structure factors.

A solution to the conflict is the real-space correlation coefficient (RSCC) plot, introduced in 1990 (ref. 4). Such plots depict, residue by residue, the fit of the model and

ligands to the electron density. Weak correlation indicates poor fit to electron density, suggesting genuine absence of ordered regions or building errors. Low real-space correlation is thus a general indication of lack of reliable information in that part of the structure.

RSCC plots for validation of deposited and released structures are available through the Electron Density Server EDS (<http://eds.bmc.uu.se/>). In addition, EDS plots are automatically generated as a part of the validation procedure during coordinate submission to the Protein Data Bank (PDB) at the European Bioinformatics Institute (EBI-MSD). Similar RSCC plots are returned by the program SFCHECK upon deposition to the US-based RCSB-PDB site.

We suggest that RSCC plots should be submitted with manuscripts to help the reviewers and users to assess the quality of structure models. Generated as a part of validation during structure deposition, these plots can be produced without any additional work by authors. The plots can be provided with the manuscript or as supplemental material to convince reviewers of the model quality in critical areas, without forcing authors to reveal coordinates and structure factors prematurely.

In view of the increasing frequency of publication of exciting and important structures of protein–ligand complexes, which depend crucially on the validity of the interpretation of ligand–protein interactions, the time is now appropriate to consider our proposal for mandatory provision of RSCC plots when submitting manuscripts to journals. Inspection of real-space correlation plots substantially lowers the risk of over-interpretation of poor local electron density, enhances information available for review, and provides authors with a strong means to demonstrate the quality of their structures without compromising their original data.

Bernhard Rupp

Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92697, USA, and q.e.d. life science discoveries, Livermore, California 94551, USA

1. *Nature* **435**, 1138 (2005).

2. *Nature Struct. Biol.* **10**, 579 (2003).

3. Kleywegt, G. J. & Jones, T. A. *Structure* **10**, 465–472 (2002).

4. Branden, C. I. & Jones, T. A. *Nature* **343**, 687–689 (1990).

Nature welcomes comments and feedback to this suggestion from researchers in the community, via authors@nature.com, or from readers at our author blog, Nautilus: http://blogs.nature.com/nautilus/2006/12/RSCC_plots.html — Editor, Nature

Contributions to Correspondence may be submitted to corres@nature.com. Published contributions are edited.

BOOKS & ARTS

Top of the pops

What's special about the best popular science books?

Jon Turney

"There is no serious or stringent idea available of what makes a book a worthy example of popularization," grumbled the British literary critic Martin Green when he tried to make sense of the science books he found in the early 1960s. Forty years later, it is still true.

The problem is even more conspicuous now, because the intervening decades have seen such an outpouring of science writing. The science publishing boom of the 1980s and 1990s was marked by blockbuster titles such as Stephen Hawking's *A Brief History of Time* (Bantam, 1988), Stephen Pinker's *The Language Instinct* (William Morrow, 1994) and Jared Diamond's *Guns, Germs and Steel* (W. W. Norton, 1997). But hundreds of less celebrated titles remain in print. And although publishers' enthusiasm has faded a little, plenty of new titles continue to appear.

As a result, the prospective reader is faced with a bewildering choice of topics, authors and treatments. We may well be in a golden age of science writing, but without some critical aids to navigation, the really good stuff may not find the readers it deserves. If some of the millions who enjoyed Bill Bryson's huge bestseller *A Brief History of Nearly Everything* (Doubleday, 2003) look for another science title, how should they choose?

Guidance is hard to find. Lists of 'best science books' typically put historic landmarks such as Newton's *Principia* or Darwin's *On the Origin of Species* alongside contemporary efforts, which is hardly helpful as they are quite different from present-day popular science books in style, audience and intent. Individual new titles tend to be reviewed in the press by enthusiasts, rather than critics. Science journalists share with fashion writers a tendency to be cheerleaders for their subject, and science book reviewers often betray an assumption that any half-way competent science popularization is a Good Thing.

So are there any qualities to look for beyond the conventional attributes of good books, such as literary style or good story-telling? Books

are complicated things, so there is no single answer. But consideration of those widely thought to be the best science writers — Jared Diamond, say, or Richard Dawkins — prompts two suggestions for what is distinctive about the best current popular science.

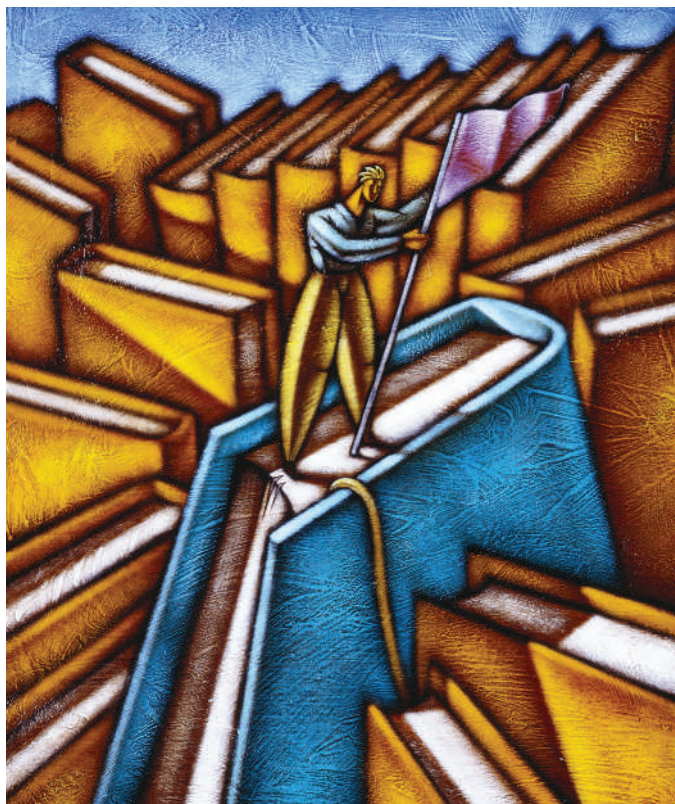
Science books often allow the authors to range beyond their own discipline. In some

range of ideas and data. But the general run of popular science books can be compared with the best on another key feature. What makes the recent boom in science writing important is that so many authors have got so good at constructing explanations. This is, of course, central to science anyway, but there is an art to recreating them on the page in intelligible form. Richard Dawkins in *Climbing Mount Improbable* (Penguin, 1996), Matt Ridley in *The Origins of Virtue* (Viking, 1996) and Brian Greene in *The Elegant Universe* (W. W. Norton, 1999) reinvent science in text of exemplary logic, clarity and accessibility. They take great pains to introduce the entities — be they genes or superstrings — that will bear the burden of the explanation, describing exactly what they can do, or what their properties and capacities are, and then showing how putting this into action produces the phenomena for which they are supposed to account.

Books still lend themselves to this better than any other medium. Building up these explanations in words often takes scores, even hundreds, of pages. The author needs great control over the different elements, and must carefully signal to the reader where they have been and where they are going. You can see all this in action in the extended exposition of the properties of genes in Dawkins' *The Blind Watchmaker* (Longman, 1986), for example, or of the elementary constituents of everything in Greene's account of string theory in *The Elegant Universe*.

Describing this kind of writing as defining the entities that underlie phenomena makes it sound rather like science. But from another point of view, the key entities in the popular science exposition become characters in an explanatory narrative. So perhaps it does all come down to story-telling. But if so, it is story-telling of a very particular kind. And it is the writers who do it best whose books will endure to become part of the popular science canon, however that is eventually defined.

Jon Turney is course leader for the MSc in creative non-fiction at Imperial College London.



hands, this creates an opportunity to forge a synthesis that offers a unique and genuine contribution to scholarship — albeit one that inevitably attracts as many brickbats as bouquets from academics more committed to their specialism. Diamond's *Guns, Germs and Steel* and his more recent *Collapse* (Viking/Allen Lane, 2005) occupy one pinnacle of this kind of writing. But there are other outstanding examples that offer multidisciplinary stimulation, such as Nick Lane's *Power, Sex, Suicide* (Oxford University Press, 2005) about the role of mitochondria in the history of life, or Stephen Mithen's *The Singing Neanderthals* (Weidenfeld & Nicolson, 2005), which speculates on the origins of music and language.

Few authors, perhaps, can command this

J. FLEMING/GETTY IMAGES

Swoop and swerve

Spinning Flight: Dynamics of Frisbees, Boomerangs, Samaras and Skipping Stones

by Ralph D. Lorenz

Springer: 2006. 346 pp. £24, \$49.95

H. K. Moffatt

What do cricket balls, frisbees, boomerangs, rotorcraft, samaras (the winged fruits of trees such as the ash or maple) and 'sensor fused weapons' all have in common? Answer: they spin as they swoop as they swerve, and are subject to a combination of aerodynamic and gyroscopic effects that determine their translational and rotational motion. In his fascinating book *Spinning Flight*, Ralph Lorenz provides a rich feast of examples of such spinning bodies, some occurring naturally, some contrived for pleasure or for a practical purpose, some exotic. He discusses the flight of each within the common framework of Euler's principles of rigid-body dynamics together with the principles of aerodynamics insofar as these determine the net force and torque experienced by the body.

Lorenz is a space scientist who worked on the control of the Huygens probe, which in early 2005 made its gyroscopic descent through the atmosphere of Saturn's moon Titan. This experience, involving the design of sophisticated on-board instrumentation, led to his wider interest in spinning flight — initially that of the ever-popular frisbee — and to recognition of the insights that might be gained from attaching miniaturized sensors to record the details of flight path and spin.

Whether in sport or in casual recreation, the side force associated with spin is a familiar phenomenon: for golfers prone to slice, it is usually a source of frustration, whereas in cricket it is exploited by bowlers to achieve deceptive swerve. This side force arises because the spin, or 'vorticity', imparted to the air in the immediate vicinity of the ball is swept into a downstream wake in the form of a pair of vortices, like those shed from the tips of aircraft wings. If backspin, say, is imparted to the ball, these vortices propagate downwards, carrying momentum with them; the associated reaction of the air on the ball is an upward force perpendicular both to the motion and the axis of spin (the Robins-Magnus effect). Although fully understood for cylindrical bodies, a rigorous derivation of this transverse force for spheres or more complex bodies is lacking, clear though the physical mechanism may be.

Children playing table-tennis soon learn how to impart topspin, backspin or sidespin to the ball, and it is here that the effect of the transverse force is most evident. In cricket, the rough seam of the ball is a deliberate complication; it triggers the transition to turbulence in the boundary layer, delaying flow separa-



Wrist action: the spin imparted on the cricket ball by bowler Shane Warne has helped Australia to lead the current Ashes series against England.

tion and leading, paradoxically, to a decrease in drag. For a spinning ball, the orientation of the seam is critical. The dimples on a golf ball reduce drag for the same reason, and so increase its range if well struck. Again, in terms of fluid dynamics, the physical effect is well understood, but unresolved problems concerning the optimal scale and shape of the dimples continue to challenge, for reasons both aesthetic and commercial. The degree to which a golf club can impart spin to the ball is a matter of impact dynamics, for which both the roughness and the compliance of the impacting surfaces are essential ingredients.

For non-spherical bodies such as a frisbee or boomerang, questions of gyroscopic stability are central. If external torques and internal dissipation (due, for example, to flexural deformation) are negligible, then, as Euler knew more than 250 years ago, such a body can spin stably about its axis of either greatest or least moment of inertia, whereas spin about the axis of intermediate inertia is unstable. The presence of aerodynamic torques modifies these conclusions, but they nevertheless provide a good starting point for a general dynamical treatment.

Lorenz covers these phenomena and their various manifestations with great skill and

economy of exposition, whetting the appetite for the more detailed treatments that he cites in an extensive list of references, many quite recent. The book is well organized, progressing from simple to more complex systems, culminating in a discussion of skipping stones (I was astonished to learn that the current record is 40 skips) and the related 'bouncing bomb' developed by Barnes Wallis and used to such effect during the Second World War — just one example of the way recreational pursuits can inspire technological development.

The discussion in the book is deliberately non-technical, and should be accessible to readers with some elementary understanding of aerodynamic principles. For the expert, the book is full of open problems that cry out for detailed investigation, whether analytical, experimental or computational. Its scope is extensive, ranging from samaras at one extreme to aircraft such as the EC-2 Hawkeye, with its rotating radar antenna, or radome, at the other. In this respect, there may be something for everyone within its attractively designed cover: a good Christmas present for the man who has everything. ■

H. K. Moffatt is in the Department of Applied Mathematics and Theoretical Physics, Wilberforce Road, Cambridge CB3 0WA, UK.

Q. ROONEY/GETTY IMAGES

Saving time

Time Restored: The Harrison Timekeepers and R. T. Gould, the Man Who Knew (Almost) Everything

by Jonathan Betts

Oxford University Press: 2006. 480 pp.
£35, \$69.50

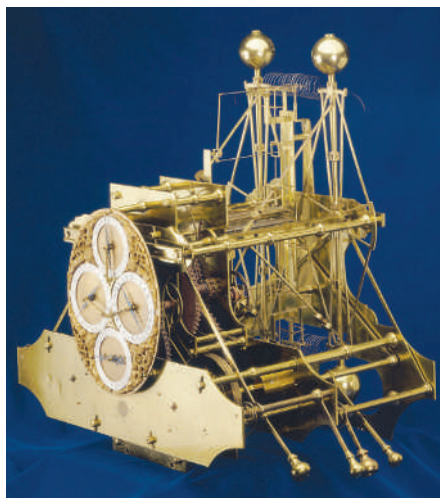
Lisa Jardine

In her best-selling popular science book *Longitude*, Dava Sobel told the tale of horologist John Harrison's eighteenth-century quest to design and build a precision timekeeper accurate enough to determine longitude at sea. Towards the end of the book, she introduces us to another 'lone genius' with a lifelong fascination with clocks, Lieutenant Commander Rupert T. Gould. It was Gould who, in the early twentieth century, found Harrison's chronometers neglected and largely forgotten in a cupboard at Greenwich Royal Observatory, where they had lain, "dirty, defective and corroded", since the 1760s. Gould devoted much of the rest of his life to restoring them to their former glory.

Without Gould's efforts, we would not today have the pleasure of seeing Harrison's timekeepers, intact and in full working order, in their own gallery in the old observatory at Greenwich. Over a period of 13 years, Gould — who had no horological training — painstakingly dismantled the timepieces, cleaned every working part, arranged for the manufacture by specialist horologists of replacement and missing pieces, and rebuilt them, meticulously documenting every step in a series of illustrated notebooks.

Jonathan Betts' compelling, information-packed new book, *Time Restored*, at last gives us a richly detailed account of Gould's life. It is a life every bit as colourful as Harrison's, and equally dogged by heartache and setback. Academically bright, although suffering from a debilitating lack of self-confidence, Gould seemed destined for a successful career in the Royal Navy. But on the eve of the First World War he suffered the first of a series of nervous breakdowns that were to blight his life. His mental health collapsed again at the outbreak of the Second World War. His periods of clinical depression were brought on by fear of three things, he said: lightning, revolution and hell. In all, he experienced four major breakdowns, during which he suffered severe memory loss and was unable to speak for long periods. Recovery each time was slow. Working with his beloved clocks was what restored him and gave him a sense of purpose.

Perhaps as a result of his mood swings, Gould's private life was a disaster. As restoring the Harrison clocks became an obsession, he neglected his children, drank heavily at home, and saw his marriage to Muriel Estall disintegrate. She left the family home, taking their



John Harrison's eighteenth-century timekeepers have been beautifully restored by Rupert T. Gould.

daughter Jocelyn with her. In the lawsuit that followed, the details of the couple's private life (she accused him of violence, drunken fits and unreasonable sexual demands) caused a scandal. Gould was ostracized by polite society and his job prospects were permanently damaged.

Gould's exceptional talents were not limited to clock restoration. He was a talented artist, producing both exquisite line drawings and bizarre imaginative art in the manner of Aubrey Beardsley. He wrote with flair and eloquence, captivating readers with a series of eccentric books on a range of curious topics, and gained a considerable reputation as a broadcaster, first as a regular contributor to the BBC's children's hour, and then as a panellist on the intellectual radio roundtable, *The Brains Trust*.

Betts has produced a finely crafted biography, full of lovingly observed insight into Gould's character, including his many personal failings. But the book is much more than a biography. In the introduction, Betts tells us how he has loved clocks since childhood, and we learn that he is now in charge of preserving the Harrison timekeepers for posterity. *Time Restored*, then, is a loving restoration of a reputation, to set alongside the clocks to which both Gould and Betts have devoted so much care and attention.

Chapter 4 is a concise retelling of the whole longitude story, including a readily comprehensible account of precisely what the problem of longitude entails, and the technical details of the timekeepers that Harrison built to solve it. Here, in one book, readers can find an authoritative account of Harrison's quest, and details of the workings of the Harrison clocks, alongside the painstaking reconstruction of the complicated life of one of the key protagonists in the history of the timekeepers.

In *Longitude*, Sobel wrote that the restored Harrison timepieces at the Greenwich old observatory "constitute John Harrison's enduring memorial, just as St Paul's Cathedral serves as monument to Christopher Wren". Betts shows us with extraordinary elegance that in the case of the Harrison clocks, the memorial is a double one: to Harrison and to Gould, the complicated, depressive, brilliant, failed navigation officer without whose obsessive attentions and unstinting labours the clocks would almost certainly have been lost for ever. ■

Lisa Jardine is director of the AHRC Centre for Editing Lives and Letters at Queen Mary, University of London, London E1 4NS, UK.

2006 wrapped up

Mary Purton

It has been a strange year for science books. Some authors have presented new ideas about science — there has been a tussle over string theory, for example, and in *Moral Minds* Marc Hauser has suggested that morality is as innate as language (see *Nature* 443, 909–910; 2006). But perhaps the dominant theme running through many of the popular science books published this year has been, surprisingly, religion.

The continuing debate about the teaching of creationism in schools has no doubt fuelled this preoccupation. Many scientists, particularly those in the United States, have been moved to take a stand against proponents of creationism and intelligent design. *Intelligent Thought*, edited by John Brockman, is a collection of essays from the likes of Jerry Coyne and

Tim White who provide elegantly expressed scientific arguments to counter the claims of intelligent design. This book should appeal to "those who already see evolutionary biology as a science", according to John Tyler Bonner (see *Nature* 442, 355–356; 2006). Michael Shermer's *Why Darwin Matters* is perhaps more accessible for the public, but neither book is likely to sway creationists from their belief.

Many of the scientists who made it to the top of the bestseller lists focused specifically on religion. Daniel Dennett's book *Breaking the Spell* provides essentially a natural history of religion but skirts around the cultural reasons why religion has developed and become such a dominant force in politics today, in the view of reviewer Michael Ruse (see *Nature* 439, 535; 2006). In *Six Impossible Things Before Breakfast*, Lewis Wolpert treads similar ground but

provides a more succinct summary of the evolution of the 'need to believe' (see *Nature* **442**, 137; 2006).

Francis Collins' *The Language of God*, an account of how his Christian faith is compatible with his work as a scientist, has sought to engage both sides in a less confrontational dialogue (see *Nature* **442**, 110 and 114–115; 2006) — as has Owen Gingerich in *God's Universe*. But Richard Dawkins isn't interested in reconciling science and religion. In *The God Delusion*, which has topped the bestseller lists in both the United States and Britain this autumn, Dawkins argues with the fervour of a preacher that religion has no place in the modern world, and that atheism is the 'true path' (see *Nature* **443**, 914–915; 2006).

Dawkins' domination of the genre of popular science books was celebrated earlier in the year with the publication by Oxford University Press of a thirtieth-anniversary edition of his book *The Selfish Gene*, and *Richard Dawkins: How A Scientist Changed the Way We Think*, a collection of comments and testimonials edited by Alan Grafen and Mark Ridley (see *Nature* **441**, 151–152; 2006).

Physicists have also been questioning our place in the Universe. Cosmologist Alex Vilenkin's *Many Worlds in One* takes a look at

the multiverse theory — the idea that many different universes exist and explanations for how we came to be in this one (see *Nature* **443**, 145–146; 2006). Paul Davies' *The Goldilocks Enigma* gives the topic a more popular treatment (see *Nature* **444**, 423–424; 2006). Playwright Michael Frayn also considers our relationship with the Universe, and much more, in his book *The Human Touch*, which will be reviewed in *Nature* next week.

After a spate of books on string theory in 2005, the hottest hope for a 'theory of everything' came in for criticism this year, with the appearance of Lee Smolin's *The Trouble with Physics* and Peter Woit's *Not Even Wrong* (see *Nature* **443**, 482, 491 and 507–508; 2006).

Jane Goodall was the subject of one of the year's notable biographies, *Jane Goodall* by Dale Peterson (see *Nature* **443**, 915; 2006). Philip Ball delved into history for *The Devil's Doctor*, his biography of Paracelsus (see *Nature* **441**, 152–153; 2006). *Broken Genius* by Joel

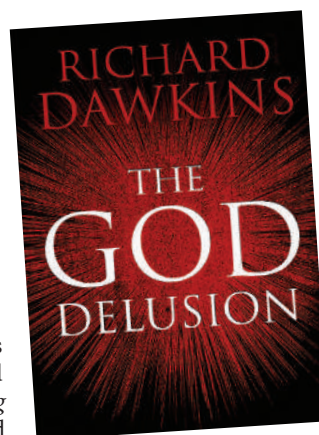
Shurkin is the first full biography of William Shockley (see *Nature* **442**, 631–632; 2006). And Francis Crick was the focus of a short biography by Matt Ridley (*Francis Crick*; see *Nature* **443**, 917–918; 2006) — a fuller treatment by Robert Olby is expected next year. This year's

Pulitzer Prize for biography, meanwhile, went to Kai Bird and Martin Sherwin for *American Prometheus*, their portrait of Robert Oppenheimer.

While many scientists have been preoccupied with religion, some novelists have turned to science for inspiration. Allegra Goodman's *Intuition* proved to be an exciting page-turner, examining the psychological motives for scientific fraud in a cancer-biology lab (see *Nature* **440**, 996–997; 2006).

Die Vermessung der Welt by

Daniel Kehlmann, a fictionalized account of scientists Alexander von Humboldt and Carl Friedrich Gauss, was a bestseller in Germany last year. An English translation by Carol Brown Janeway, *Measuring the World*, is now available and will be reviewed in *Nature* next month. ■ Mary Purton is *Nature's* book review editor.



CHRISTMAS READING

A selection of books on the lighter side of science for the holiday period.

Why Don't Penguins' Feet Freeze?

edited by Mick O'Hare (Profile, £7.99).

The book version of *New Scientist's* 'Last Word' column continues to top the bestseller lists in the run-up to Christmas with more than 180,000 copies sold in Britain so far.

Giant Leaps

by John Perry & Jack Challoner (Boxtree, £12.99)

This is an amusing guide to the key discoveries, inventions and events in science, technology and medicine from UK newspaper *The Sun* and London's Science Museum, told in 'sensational' tabloid style (see picture).

Fly

by Steven Connor (Reaktion Books, £12.95, \$19.95)

Perhaps the perfect

present for a *Drosophila* geneticist, this book is an exploration of flies through myth, literature, art and biology.

How to Cut a Cake

by Ian Stewart (Oxford University Press, £9.99, \$14.95)

Various mathematical conundrums are featured, such as why phone cords get tangled, which way of tying shoelaces uses the shortest amount of lace, and how to play never-ending chess.

and tobogganing otters. Both have a bibliography but no index.

Bang!

by Brian May, Patrick Moore & Chris Lintott (Carlton, £20)

Queen guitarist and astrophysicist May joins the presenters of *The Sky At Night* for this lavishly illustrated "complete history of the Universe".

Ken Libbrecht's Field Guide to Snowflakes

by Ken Libbrecht (Voyageur Press, \$12.95, £7)

Explore the icy world of snowflakes, from stellar dendrites to sector plates, with this wonder of microphotography.

For more serious reading, Oxford University Press has themed box sets from its Very Short Introductions series. **The Brain Box** (£25) has books on evolution (by Brian and Deborah Charlesworth), consciousness (Susan Blackmore), intelligence (Ian Deary), cosmology (Peter Coles) and quantum theory (John Polkinghorne).



Moths That Drink Elephants' Tears

by Matt Walker (Portrait Books, £9.99)

Why Pandas Do Handstands

by Augustus Brown (Bantam Press, £9.99)

These two are collections of facts about curious animal behaviour. So if you want to know which cats purr and which don't, or why female brown trout fake orgasms, Walker's book is the one. Brown offers similar fare, including winking cuttlefish

NEWS & VIEWS

OPTICS

Momentum in an uncertain light

Ulf Leonhardt

How much momentum does light transfer to a material through which it passes? This is a surprisingly opaque matter, contested for almost a century, that is still the object of theory and experimentation.

One of Rudolf Peierls' *Surprises in Theoretical Physics*¹ is the difficulty of assessing the momentum of light in transparent materials such as glass or water. Despite its being a seemingly basic concept, rival theories^{2,3} predict values for the momentum that differ by considerable factors, and that have been hotly debated for many decades. Experimental tests^{4,5} have been rare. Two papers now bring out the big guns to break the impasse: writing in *Physics Letters A*, Dereli, Gratus and Tucker⁶ resort to Einstein's general relativity for the requisite theory; and in *Physical Review Letters*, Campbell *et al.*⁷ use ultracold atoms in an experimental test.

Light's momentum describes the degree to which light sets other things in motion when they absorb or reflect it. The force of light is usually rather weak, but visual evidence for its existence can be seen, for example, in a comet's tail. Here, the action of sunlight transfers momentum onto the particles of the tail, pushing dust away from the comet (Fig. 1).

The debated issue is the ratio between the momentum, p , and the energy, E , of light in a medium. This ratio depends on the degree to which a material slows light from its speed in a vacuum, c . The reduced speed is often written c/n , where n is the refractive index of the medium. In 1908, Hermann Minkowski² proposed that the fundamental relationship between all these quantities is $p = nE/c$; a year later, Max Abraham³ postulated $p = E/(nc)$ (Box 1, overleaf). These rivaling momenta differ by n^2 , which is a sizeable factor in most media: in water, n alone is 1.33, and in glass it is 1.46.

It might seem that this dispute could easily be settled by experiment: all it should take is some water and a laser beam illuminating its surface. The surface partially reflects the beam; the rest of the incident light enters the water. The water must take up any imbalance in momentum, so the surface should rise or fall, depending on the momentum of light in water compared with that in air. In Minkowski's case, the



Figure 1 | Light, momentum, action. The comet Hale-Bopp, seen here over the Joshua Tree National Park in southern California on the evening of 28 March 1997, has both a blue ion tail and a white dust tail. Whereas the ion tail is carried away by the 'solar wind' of charged particles from the Sun's atmosphere, the dust tail is pushed by the radiation pressure of the sunlight. The momentum transfer in this second case is weaker than that in the first, resulting in the splitting of the tails.

momentum in water is higher and the water should rise; in Abraham's case, the reverse is true and the water level should fall⁸. An experiment in the mid-1970s found that the water rises⁴. In a second experiment, the light pressure on a mirror suspended in water was measured⁵, and the same result emerged. Thus, in both cases, Minkowski is the winner. Case closed.

Not so fast. Sometimes, one can directly calculate light forces without considering the momentum transfer. Careful calculation⁸ of the forces acting behind the scenes reveals that the momentum of light is in fact chameleonic: its precise value depends on the type of experiment performed. In fact, according to the calculations, the first of the experiments⁴ should have observed Abraham's momentum in action, whereas the second⁵ should indeed

see Minkowski's momentum. The problem with the first experiment was that the beam did not uniformly illuminate the surface. This imbalance created lateral light forces that curved and lifted the water to a greater extent than the momentum could push it, resulting in a false interpretation of the result.

Campbell *et al.*⁷ applied for the first time the sophisticated tools of atom optics to measuring light's momentum transfer. Like light, atoms are simultaneously particles and waves. In traditional optics, instruments made of matter act on light. In atom optics, the roles are reversed: light acts on atom waves. Campbell and colleagues performed their experiments on a few million atoms cooled to nearly absolute zero. At such low temperatures, atom waves oscillate in unison. The authors illuminated the atoms with two beams of light travelling in opposite directions, which together form a so-called standing wave. When the atoms absorb the photons, they begin to move in the two directions of the light, and the atom waves interfere with each other. The momentum transfer could be inferred from the interference fringes by fitting the profile of the fringes to the theory.

To the surprise of the leader of the experimental group⁷, Wolfgang Ketterle, it fitted Minkowski's prediction.

Theory can use equally sophisticated tools as experiment: for example, the general theory of relativity used by Dereli, Gratus and Tucker⁶. According to this theory, mass and momentum — including those of light — curve space and time. This effect is felt as the force of gravity. The authors first obtained Abraham's momentum in a similar way to that detailed in a long-forgotten paper from 1923 (ref. 9). This exploited an intriguing idea that originates from Fermat's 'principle of least time': that, because light rays choose their paths to minimize their travelling time, they tend to stay as long as possible in regions where the refractive index n is low and speed c/n high, and avoid regions of high n . In other words,

W. PACHOLKA/LONG BEACH PRESS/NEWS.COM

Box 1 | Theoretical routes to light's momentum

The reasoning behind both the Minkowski and the Abraham formulae for the momentum of light is best explained with a little support from two pillars of modern physics: wave-particle duality and Einstein's relativity.

Wave-particle duality is a fundamental tenet of quantum mechanics, and states that light, like everything else in the quantum world, simultaneously has characteristics of both a particle and a wave. Light acting as a wave oscillates

with period τ and wavelength λ . During each period, the wave advances by an amount determined by its speed in the medium that contains it: thus, $\lambda = (c/n)\tau$, where c is the speed of light *in vacuo* and n is the refractive index of the medium. Light acting as particles consists of photons, each of which carries an amount of energy $E = h\nu$, where h is Planck's constant. Combining these two expressions, and after a little rearrangement, you obtain Minkowski's prediction² for

light's momentum, $p = nE/c$, if you require $p = h/\lambda$. This fundamental relationship between momentum and wavelength now bears the name of the de Broglie relation.

Alternatively, however, you can start from Einstein's energy-mass equivalence formula $E = mc^2$. This implies that light has mass, and mass multiplied by velocity c/n gives momentum. The expression that emerges in this case is $p = E/(nc)$, Abraham's momentum³.

U.L.

the refractive index defines an optimization measure in space and time: a material acts as a space-time geometry of its own⁹.

Dereli, Gratus and Tucker⁶ also found a way to derive the Minkowski momentum from the principles of general relativity. The fact that they could deduce expressions for both momenta from the same starting point might be connected to another twist of the story. As atom optics has shown, the role of light and matter can be reversed. So, if a transparent material — the glass of a lens, for example — acts on light as though to change the geometry of a situation, light should change the geometry of such materials as well. This leads to a theory of the momentum of light in ultracold atoms¹⁰: whenever the wave aspects of atoms dominate, as in Campbell and colleagues' interference experiment⁷, the Minkowski momentum appears, but when the particle aspects are probed, the Abraham momentum is relevant.

How this might apply to the momentum of light in more ordinary materials such as glass

or water remains unclear. When exactly is the Minkowski momentum applicable, and when the Abraham momentum? Whatever the final answer, one thing is clear: light continues to surprise.

Ulf Leonhardt is in the School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, UK.
e-mail: ulf@st-andrews.ac.uk

1. Peierls, R. *More Surprises in Theoretical Physics* (Princeton Univ. Press, 1991).
2. Minkowski, H. *Nachr. Ges. Wiss. Göttin Math.-Phys. Kl.* **53** (1908).
3. Abraham, M. *Rend. Circ. Matem. Palermo* **28**, 1 (1909).
4. Askin, A. & Dziedzic, J. M. *Phys. Rev. Lett.* **30**, 139–142 (1973).
5. Jones, R. V. & Leslie, B. *Proc. R. Soc. Lond. A* **360**, 347–363 (1978).
6. Dereli, T., Gratus, J. & Tucker, R. W. *Phys. Lett. A* **361**, 190–193 (2007); preprint available at <http://arxiv.org/math-ph/0610078> (2006).
7. Campbell, G. K. *et al. Phys. Rev. Lett.* **94**, 170403 (2005).
8. Loudon, R. *J. Mod. Optics* **49**, 821–838 (2002).
9. Gordon, W. *Ann. Phys. (Leipz.)* **72**, 421–438 (1923).
10. Leonhardt, U. *Phys. Rev. A* **73**, 032108 (2006).

MALARIA

A protective paradox

Stephen L. Hoffman

The infectious form of the malaria parasite has thousands of proteins, making it tough to develop a vaccine for it. Narrowing down which proteins cause protective immune responses may help resolve the problem.

A vaccine against malaria would be the ideal means of preventing the hundreds of millions of cases of the disease that occur annually across the globe¹. But no malaria vaccine has yet been licensed, and there is little consensus on how to develop one. Initial success in inducing an immune response against the malaria parasite (*Plasmodium falciparum*) came more than 30 years ago^{2–4}, with the discovery that human volunteers could be completely protected against malaria by exposure to radiation-weak-

ened sporozoites — the infectious parasite cells passed on by mosquitoes (Fig. 1). But until recently it has not been considered practicable to manufacture and administer such 'attenuated' sporozoites as part of a mass immunization strategy⁵ — not least because it took more than 1,000 bites from mosquitoes infected with the irradiated cells to fully protect most of the volunteers against the disease⁴.

Considerable effort has therefore gone into trying to define which of the thousands of

proteins expressed by the parasite^{6,7} are involved in the protective immunity elicited by the attenuated sporozoites, with the aim of eventually making vaccines against them. On page 937 of this issue, Kumar *et al.*⁸ report that one of these proteins, the circumsporozoite protein, might elicit much of the protective immunity induced by the attenuated sporozoites⁴. Puzzlingly, however, it seems that a protective response can occur in the absence of this protein too.

The first clinical trials using a purified-protein malaria vaccine began 20 years ago^{9,10}, but so far only volunteers immunized with vaccines based on the circumsporozoite protein¹¹ — the major sporozoite surface protein — have been reproducibly protected against *P. falciparum* sporozoite challenges. However, the protective immunity induced by the best circumsporozoite-protein vaccine is far lower¹² than that induced by radiation-attenuated *P. falciparum* sporozoites⁴. One explanation for this dramatic difference is that the whole-parasite vaccine induces protective immune responses against tens, hundreds or even thousands of parasite proteins, whereas the single-protein vaccine can elicit an immune response only against that protein¹³.

Accordingly, there are significant efforts to identify additional target proteins expressed in sporozoites, and in the parasite during the liver stages of infection, to construct a more effective vaccine. If one, or a few proteins, like the circumsporozoite protein, is responsible for the protection, efforts to identify other target proteins could be curtailed, and resources focused on optimizing immunization with these few proteins. If, however, strong immunity depends on immune responses against many proteins, it may be difficult, or even impossible, to construct an effective vaccine based on only parts of the parasite. So, resolution of whether protective immunity relies on one, a few or many proteins is crucial.

Kumar *et al.*⁸ used a mouse model of malaria where the animals are infected with the *Plasmodium yoelii* parasite, as *P. falciparum* does not produce malaria in mice. This mouse model is meaningful for human malaria because experiments using attenuated *P. yoelii* sporozoite vaccines have consistently paralleled or predicted results in humans exposed to attenuated *P. falciparum* sporozoites. To examine how important the circumsporozoite protein is in inducing immunity to attenuated sporozoites, the authors first genetically engineered mice that could not generate an immune response to the *P. yoelii* circumsporozoite protein (PyCSP).

Mammalian immunity is based on two broad types of immune reaction — antibody-based immunity and the cell-based immunity that involves T cells. Kumar *et al.* therefore had to disrupt both of these reactions to create PyCSP-tolerant mice. They first injected the gene

*This article and the paper concerned⁸ were published online on 6 December 2006.

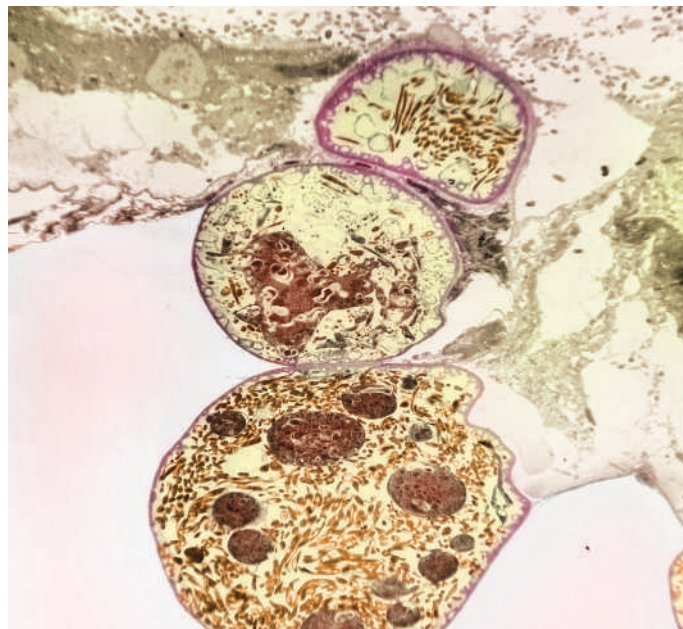


Figure 1 | Malaria parasites. If a mosquito feeds on malaria-infected blood, the parasite attaches itself to the insect's gut wall and forms 'oocysts'. These spherical structures support the development of sporozoites (small, brown shapes), which can be transferred back to humans through the mosquito's saliva.

encoding PyCSP into mouse oocytes (eggs). The resulting transgenic mice produced PyCSP as though it were a normal mouse protein. This meant that PyCSP should not be recognized as potentially harmful foreign material, so the mice would not raise a T-cell immune response to the protein. The transgenic mice were then bred with mice that were unable to produce any antibodies and thus had no antibody-based immunity against the PyCSP protein either. The final mice were therefore expected to be immunologically tolerant to PyCSP. Scientists have been attempting to engineer such mice for more than 15 years, so these animals are a considerable technological achievement by Kumar and colleagues.

The authors immunized the PyCSP-tolerant mice with attenuated *P. yoelii* sporozoites to determine whether animals that could not respond to PyCSP could still produce protective immunity against the malaria parasite. When the mice were immunized with two doses of attenuated *P. yoelii* sporozoites and then infected with untreated sporozoites, they produced only minimal protection — there was only a 10% reduction in the number of copies of parasite ribosomal RNA in the animals' livers compared with non-immunized controls. In contrast, normal mice that had been immunized had a 10,000-fold reduction in copy number compared with controls. So, lack of an immune response to PyCSP led to a dramatic reduction in protection as measured by parasite burden in the liver.

However, immunization with two doses of attenuated *P. yoelii* sporozoites often does not fully protect mice, whereas immunization with three doses is almost always associated with 100% protection against development of blood-stage infection — the gold standard of protection against malaria. When Kumar *et al.* immunized their PyCSP-tolerant mice with three doses of attenuated *P. yoelii* sporozoites, and then infected them with normal *P. yoelii*

sporozoites, there was 100% protection against blood-stage infection.

Therefore, PyCSP can elicit an early, if incomplete, protective immune response. In fact, the authors conclude that PyCSP is responsible for 90% of the protection, and is thus immunodominant — that is, the majority of the protective cellular and antibody responses are directed against this protein rather than against other proteins. However, the experiment with three doses of attenuated sporozoites demonstrates that an immune response against PyCSP is not required for protective immunity.

Although these results strongly suggest that PyCSP is involved in the protection elicited by radiation-attenuated *P. yoelii* sporozoites, they do not definitively establish this because several possible alternative interpretations exist. The most obvious explanation for the difference in protection in the transgenic mice after two and three doses of *P. yoelii* sporozoites is that it took the full three doses to elicit a protective immune response against one or many proteins other than PyCSP, and that these other proteins — not PyCSP — elicited the whole protective immune response. In normal mice, these proteins could be the immunodominant ones, or at least equal partners in the full protection that is produced by three doses.

Another explanation concerns the fact that there are two types of T cell: effector T cells that actually do the job of eliminating invading pathogen-infected cells, and helper T cells that facilitate the work of the effector T cells. So it may be that responses against proteins other than PyCSP were responsible for effector-T-cell-induced protective immunity after two or three doses of the vaccine, but that helper-T-cell responses against PyCSP were required to facilitate protective T-cell responses against other proteins after two doses, but not after three.

A third possibility is that PyCSP truly is the immunodominant protective antigen in



50 YEARS AGO

The U.S. National Academy of Sciences at its autumn meeting in Washington adopted the following resolution respecting recent events in Hungary: "[We]... unite in expressing [our] profound admiration and sympathy to fellow scientists in Hungary and to all the men and women of that nation who have demonstrated their love of liberty with sacrificial devotion during the tragic events of the past few weeks. American scientists look forward with hope to a time when their Hungarian colleagues, freed from external oppression, will be able to join fully in the international exchange of information, discussion and encouragement which is essential to the progress of science".

From *Nature* 15 December 1956.

100 YEARS AGO

There is much knowledge enshrined in Parliamentary Blue-books, and doubtless some wisdom. Very often it remains enshrined in them. A better fate, however, has awaited the report of the Departmental Committee appointed by the President of the Local Government Board in 1899 "to inquire into the use of preservatives and colouring-matters in foodstuffs."... The policy of our laws has been to allow food-producers a free hand, subject to the restriction that any preservative added shall not render the food injurious to health. But has this *laissez faire* attitude been a wise one? True, it leaves the food manufacturer free to experiment—which is, so far, so good. But it gives him the consumer's living body as *corpus vile*—which is not so good... Let a responsible body be appointed, competent to examine the newer substances; let it hear what is to be said on either side, and let it make whatever experiments are necessary and practicable to test the evidence. And let no preservative or colouring-matter whatever be added to foodstuffs until it has been at least provisionally approved by this responsible authority.

From *Nature* 13 December 1906.

50 & 100 YEARS AGO

attenuated *P. yoelii* sporozoites, and that mice were tolerant to PyCSP after two doses of the vaccine but that the third dose somehow overcame the tolerance to PyCSP-mediated protection. This is unlikely, but as no T-cell assays were reported, confirmation or refutation of this explanation awaits additional experiments.

Kumar and colleagues have significantly advanced our understanding of protective immunity in malaria. However, their results demonstrate the difficulty of determining the relative importance of a single protein in the protective immunity elicited by a vaccine consisting of a live, attenuated, whole, infectious agent. The results also underscore the incompleteness of our understanding of immunodominance and subdominance¹⁴. Nonetheless, these PyCSP-tolerant mice are a unique resource that will allow scientists to clarify the role of PyCSP in protective immunity, and perhaps further the development of a highly effective vaccine that will reduce the misery wreaked by malaria. ■

Stephen L. Hoffman is at Sanaria Inc., 12115 Parklawn Drive, Suite L, Rockville, Maryland 20852, USA.
e-mail: slhoffman@sanaria.com

1. Breman, J. G., Egan, A. & Keusch, G. T. *Am. J. Trop. Med. Hyg.* **64** (Suppl. 1–2), iv–vii (2001).
2. Clyde, D. F., Most, H., McCarthy, V. C. & Vanderberg, J. P. *Am. J. Med. Sci.* **266**, 169–177 (1973).
3. Rieckmann, K. H., Carson, P. E., Beaudoin, R. L., Cassells, J. S. & Sell, K. W. *Trans. R. Soc. Trop. Med. Hyg.* **68**, 258–259 (1974).
4. Hoffman, S. L. *et al. J. Infect. Dis.* **185**, 1155–1164 (2002).
5. Luke, T. C. & Hoffman, S. L. *J. Exp. Biol.* **206**, 3803–3808 (2003).
6. Gardner, M. J. *et al. Nature* **419**, 498–511 (2002).
7. Hoffman, S. L., Subramanian, G. M., Collins, F. H. & Venter, J. C. *Nature* **415**, 702–709 (2002).
8. Kumar, K. A. *et al. Nature* **444**, 937–940 (2006).
9. Ballou, W. R. *et al. Lancet* **1** (8545), 1277–1281 (1987).
10. Herrington, D. A. *et al. Nature* **328**, 257–259 (1987).
11. Dame, J. B. *et al. Science* **225**, 593–599 (1984).
12. Kester, K. E. *et al. J. Infect. Dis.* **183**, 640–647 (2001).
13. Hoffman, S. L., Rogers, W. O., Carucci, D. J. & Venter, J. C. *Nature Med.* **4**, 1351–1353 (1998).
14. Holtappels, R. B. I. *F. Futura* **20**, 157–163 (2005).

Potential competing financial interests: declared; details accompany the article on *Nature's* website.

LEUKAEMIA

Niche retreats for stem cells

David A. Williams and Jose A. Cancelas

Leukaemic cells and normal blood-producing cells relate differently to their surroundings. This concept has now been extended to leukaemic stem cells, suggesting a fresh approach to therapy.

Maintaining good neighbourhood relations is important for all cells in multicellular organisms. The complex interactions between adjacent cells and with their local microenvironment support the growth, development and function of the cells — and cancer cells are no exception. Writing in *Nature Medicine*, Jin *et al.*¹ and Krause *et al.*² report convincing evidence from two different models that CD44, a cell-surface receptor that mediates cell–cell contacts, helps the cells that give rise to leukaemia to engraft in the bone marrow. Notably, the normal cells that give rise to blood in the bone marrow do not apparently rely to the same extent on this receptor for their function, suggesting that CD44 might be a therapeutic target for leukaemia.

Blood cells develop from immature cells known as haematopoietic stem cells and progenitor cells (HSC/Ps). In addition to being programmed with the capability to make blood cells, these HSC/Ps also divide to maintain their own numbers, a process termed self-renewal. HSC/Ps have a defined spatial organization in the bone-marrow cavity, with the most-primitive cells being located in stem-cell niches³ near the endosteum of the bone — the layer of connective tissue that lines the cavity. But HSC/Ps don't exist in isolation; they prefer company.

For instance, they thrive in culture only when they are in contact with stromal cells derived from the bone marrow⁴. In the body, the endosteal niche may also include osteoblast cells, which are responsible for bone formation, and it seems that these cells too may have a direct role in HSC/P function^{5,6}, but the molecular pathways involved still need to be completely defined.

At least some HSC/Ps may also reside in a vascular niche located in the vascular network of the bone marrow (the sinusoids)^{7,8}. In each of these stem-cell niches, cell-surface receptors expressed on HSC/Ps, including CD44, c-kit, CXCR4 and β 1 integrins, seem to be vital for the localization and retention in the bone marrow.

The HSC/Ps that are circulating in the blood and engraft into the marrow may gain entry into the bone-marrow cavity in a manner somewhat analogous to the better-understood process by which white blood cells exit blood vessels to travel to sites of inflammation. This involves first rolling and becoming firmly attached to the vessel wall, and then traversing this endothelial boundary and migrating into the complex cellular milieu of the bone-marrow cavity. Ultimately, these cells settle into a haematopoietic microenvironment or

stem-cell niche (Fig. 1, overleaf). Interrupting this migration process prevents engraftment or retention of normal HSC/Ps in the bone marrow, stopping these cells from self-renewing and contributing to blood formation. It is not yet known whether circulating HSC/Ps play any physiological role in normal haematopoiesis, or whether they are merely travelling from one spot in the bone marrow to take up residence in another.

So how does this ordered scheme go awry in leukaemia? The current view of how this cancer arises posits that there are leukaemia stem cells (LSCs) that share characteristics with HSC/Ps, but that initiate and maintain leukaemia instead of normal blood-cell development⁹. However, the nature of LSC interaction with the supporting bone-marrow microenvironment, where leukaemia presumably arises, is unclear.

Both Jin *et al.*¹ and Krause *et al.*² based their work on the idea that — similar to normal HSC/Ps — LSCs depend upon interactions within a specific niche, the LSC niche. Notably, the investigators hypothesized that some of these interactions are specific for LSCs alone and therefore could be a target for therapy. In both reports, different forms of CD44 seem to be crucial components in the interaction of LSCs with a supportive haematopoietic microenvironment. Both papers show that altering the function of CD44 leads to delay in the progression of leukaemia in mouse models of the disease.

CD44 is a multifunctional, ubiquitously expressed protein that helps cells to adhere to other cells and to matrix proteins, and it participates in the recruitment of certain white blood cells to sites of inflammation and in their migration through lymphatic tissues. In HSC/Ps, the isoforms of CD44 bind to its ligand, hyaluronic acid, which is expressed in the bone-marrow sinusoids and in the endosteal region. This binding seems to be specifically important for allowing LSCs to settle in the bone marrow and for subsequent leukaemia development (Fig. 1)^{1,2}.

Jin *et al.*¹ used samples from patients suffering from acute myeloid leukaemia to seed tumours in mice that lack an immune system, so that the animals would not reject the human cells. The authors report that treatment with an activating antibody directed against CD44 — which would be expected to block any interactions of CD44 with other proteins — significantly reduced tumour burden and prolonged survival of the leukaemia-engrafted mice.

Mechanistic studies suggest that dual molecular pathways are involved in the effect of CD44 inhibition on leukaemia progression. Antibody treatment of the putative LSCs from the human patients differentially inhibited the migration of these cells to the bone marrow and spleen. Antibody treatment also directly altered the fate of LSCs by inducing development into a more mature state. The relevance of this latter finding *in vivo* was suggested by the direct

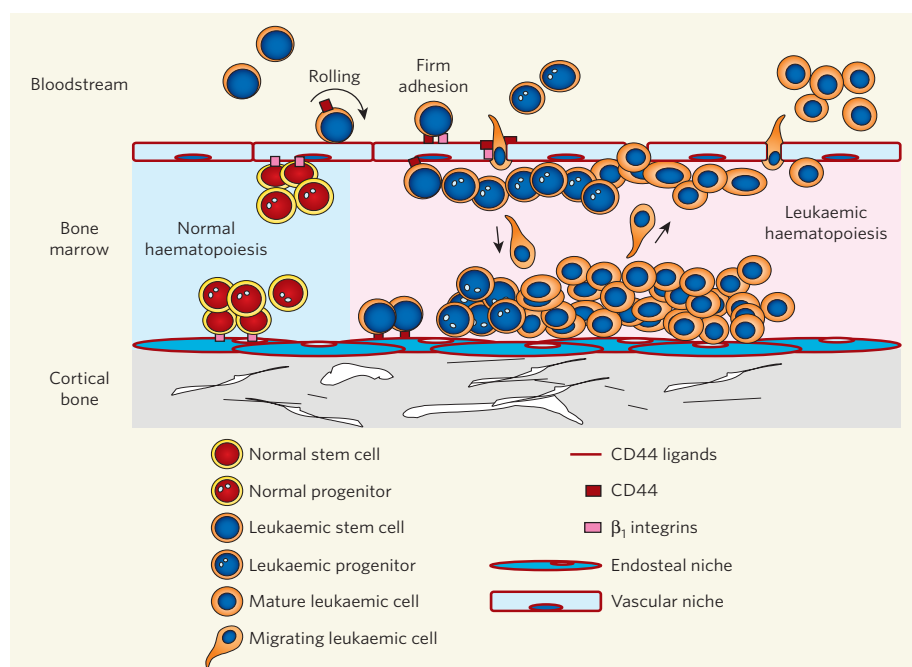


Figure 1 | Putative migration of stem cells between the bloodstream and bone marrow. Normal stem cells require multiple receptors to adhere firmly to the endothelium (the cell layer that lines blood vessels). They then make their way to putative endosteal and vascular stem-cell niches in the bone marrow. The β_1 integrins are shown as an example of the adhesion molecules involved in migration and adhesion of normal stem cells to their niches. Jin *et al.*¹ and Krause *et al.*² show that leukaemia stem cells instead seem to depend on the expression of CD44 to localize to the bone marrow and initiate a leukaemia-like disease in mouse models.

injection of antibody-treated human cells into the mouse marrow cavity. Such injected cells failed to establish a robust leukaemia even though these cells were not required to traverse the usual engraftment pathways.

Krause *et al.*² used a mouse model of chronic myelogenous leukaemia, which is caused by the expression of the Bcr–Abl fusion protein. They report similar — but not identical — results. To examine the role of CD44, the authors expressed Bcr–Abl in mouse HSC/Ps that were genetically deficient in CD44, and transplanted these cells into recipient mice of the same strain as the donors (to avoid rejection). In these studies, the progression of Bcr–Abl-induced, leukaemia-like disease was delayed in recipient mice. In addition, similar to the results of Jin *et al.*¹, antibody treatment of recipients of Bcr–Abl-expressing LSCs also slowed disease progression. However, unlike the studies with human leukaemia cells, the direct injection of antibody-treated LSCs led to disease similar to that in control-treated cells.

The basis for this difference in the two studies is not clear, but it could be related to the specific tumour type (acute versus chronic), model (cross-species versus same-strain transplant) or antibody used. This difference may be quite important because it suggests that at least one mechanism noted above — the direct effect of CD44 binding to its ligand on LSC differentiation — is potentially not occurring in both models. Additional studies are clearly needed.

The different effect of antibody treatment on LSCs and normal HSC/Ps is intriguing

and important, but the basis for the difference is not yet clearly defined. Work by Jin *et al.*¹ suggests this difference may be because different isoforms of CD44 are expressed by LSCs compared with HSC/Ps. Krause *et al.*²

suggest instead that the differential effect may relate to the relative level of CD44 expression in the leukaemia-initiating cells. Similar issues have plagued past studies implicating CD44 in the homing and engraftment of normal HSC/Ps into the bone marrow. In either case, the results are not yet definitive. In spite of this continued uncertainty, these reports^{1,2} provide significant insight into the biology of leukaemia-initiating cells with stem-cell-like characteristics and normal HSC/Ps with respect to interaction with the haematopoietic microenvironment. Such differences in the behaviour of the normal and leukaemic cells could well provide new avenues for targeted therapies.

David A. Williams is in the Division of Experimental Hematology, Cincinnati Children's Research Foundation, Cincinnati, Ohio 45229, USA. Jose A. Cancelas is in the Hoxworth Blood Center, University of Cincinnati Medical Center, Cincinnati, Ohio 45267, USA.
e-mail: David.Williams@cchmc.org

1. Jin, L., Hope, K. J., Zhai, Q., Smadja-Joffe, F. & Dick, J. E. *Nature Med.* **12**, 1167–1174 (2006).
2. Krause, D. S., Lazarides, K., von Andrian, U. H. & Van Etten, R. A. *Nature Med.* **12**, 1175–1180 (2006).
3. Lord, B. I., Testa, A. G. & Hendry, J. H. *Blood* **46**, 65–72 (1975).
4. Dexter, T. M. *et al.* *J. Cell Physiol.* **91**, 335–344 (1976).
5. Nilsson, S. K., Johnston, H. M. & Coverdale, J. A. *Blood* **97**, 2293–2299 (2001).
6. Calvi, L. M. *et al.* *Nature* **425**, 841–846 (2003).
7. Kiel, M. J., Yilmaz, O. H., Iwashita, T., Terhorst, C. & Morrison, S. J. *Cell* **121**, 1109–1121 (2005).
8. Cancelas, J. A. *et al.* *Nature Med.* **11**, 886–891 (2005).
9. Lapidot, T. *et al.* *Nature* **367**, 645–648 (1994).

SEMICONDUCTOR ELECTRONICS

Organic crystals at large

Paul Heremans

Fabricating large-scale semiconducting surfaces for the flexible screens of the future is a bothersome business. A simple technique for growing single-crystal organic semiconductors brings new vision to the field.

On page 913 of this issue, Zhenan Bao and colleagues (Briseno *et al.*)¹ describe a method for nucleating and growing single-crystal organic semiconductors over electrodes on an arbitrary substrate, and so producing extended arrays of high-performance electronic transistors. Their method represents a step towards practical applications of high-quality, but fragile, single crystals of organic semiconductors, and could show the way to high-performance electronic devices that extend over large and flexible surfaces.

Transistors are minuscule switches that lie at the heart of all electronic devices — computers, memory sticks, displays, you name it. At the heart of every transistor is a channel of semiconductor material that bridges two

electrodes. An electrical potential applied at a third electrode, the 'gate', controls whether current is allowed over this bridge or not. The most successful semiconductor for such switching applications, omnipresent in today's electronics, is silicon.

Silicon transistors can be processed on two very different types of substrate. The first is a thin, single-crystal, nearly defect-free silicon substrate called a wafer. Wafers are limited in size, but can be processed in sophisticated ways to bear a mind-boggling density of nanoscale transistors: a wafer 300 millimetres in diameter can hold hundreds or thousands of chips, each one containing up to a billion transistors. Wafer-based technologies are continuously scaling to larger transistor counts and density

CONSERVATION BIOLOGY

Unkind cuts for incense

Gold, frankincense and myrrh — the three royal gifts of the Christmas story — remain valuable commodities. But as Toon Rijkers *et al.* report in the *Journal of Applied Ecology* (43, 1188–1195; 2006), the latter-day story of frankincense is also a tale for our times.

Frankincense is a resin produced by several small trees of the genus *Boswellia*, which grow in sub-Saharan Africa, from Nigeria to the Horn of Africa, and Arabia and the Himalayan foothills. The resin is exuded naturally from leaves and shoots, but especially from wounds. Its natural function is probably to ward off grazers and fungal attack. But the resin's fragrance has also made it prized as a perfume, fumigant, flavouring and medicinal compound.

Rijkers *et al.* looked at frankincense production in Eritrea, where it is used locally for cultural and medicinal

purposes, and is also exported. Wild *Boswellia papyrifera* trees are tapped by making incisions around the trunk of the tree, starting in mid-September at the end of the summer monsoon. The resin is harvested every three weeks by reopening the incisions, and harvesting continues throughout the dry season. *Boswellia* trees also produce their flower buds at the end of the wet season, however. Could the constant harvest of resin, commencing at the time of floral induction, be a serious drain on the trees' carbon resources? Not least because leaves are lost at the beginning of the dry season, so carbon for flowering and resin production must be obtained from storage products.

To investigate the effects of resin-tapping on the tree life-cycle, and especially seed quality, Rijkers *et al.* checked the seed-germination potential of tapped and untapped



trees. They found that seeds from annually tapped trees had a germination success of only about 20% compared with a figure of 80–90% for untapped trees. Tapped trees also produced fewer inflorescences, fruits and seeds. So tapping is evidently costly to the individual tree.

There has been a severe reduction in the natural regeneration of

Boswellia in Eritrea, to which more intense tapping may be contributing. Actions that would help to reverse this trend include reducing the amount of resin taken per tree, and resting selected trees for a few seasons; the *Boswellia* populations also need protection from intense grazing.

Here, however, the spectre of the 'Tragedy of the Commons' arises: the resources for both frankincense harvesting and grazing are freely available, and so prone to overexploitation unless the common good is taken into account. Self-regulation in the cause of long-term sustainability is a difficult lesson to learn, as a wider world has found out from the over-exploitation of fisheries.

Peter D. Moore

Peter D. Moore is in the Department of Biochemistry, King's College London, Franklin-Wilkins Building, London SE1 9NH, UK. e-mail: peter.moore@kcl.ac.uk

C. OSBORNE/CORBIS

per chip, allowing ever increasing computing power and ever higher memory density.

But other applications require transistors to be sparsely distributed over large areas, and can't afford to be limited by wafer size. A flat-panel display, for example, has transistors distributed over its whole surface that switch the miniature pixels forming the image on and off. Larger screens require the distribution of transistors over larger areas, and using single-crystal semiconductor wafers as substrates is simply not economical in this case. Instead, large-area applications use 'foreign', non-semiconductor substrates, such as metals, ceramics or glass, that can be manufactured as large sheets. In the future, substrates might also include flexible and less brittle materials such as plastic foils².

Fabricating transistors with sufficient performance and high yield over large areas on foreign substrates is enormously challenging. Most conventional semiconductors, including silicon, can be forced to grow as defect-free films only if the growth substrate provides a perfect template for their crystal structure. When deposited on an arbitrary substrate, they tend to form amorphous, defect-infested films. Decades of research have been invested in improving the quality of these amorphous semiconductors to levels acceptable for the display industry.

One particular improvement involves recrystallizing the amorphous film into a polycrystalline film consisting of small, single-crystal grains, each with a particular crystal orientation, separated by thin boundary regions. The

properties of polycrystalline materials are less ideal than those of monocrystalline materials, but are in many respects better than those of amorphous materials. But the recrystallization process requires high temperatures, so, although it can be applied with success to, say, a glass substrate, it is delicate and difficult to scale the process up to an industrial level for substrates such as plastic foil that cannot withstand elevated temperatures.

Because the transistors in large-area applications are distributed over the surface, a high-quality semiconductor film is in fact not required over the entire area. A viable way around the large-area problem could thus consist of providing semiconductor material only at those places that have to bear transistors.

Briseno *et al.*¹ come up with an original solution to achieve precisely this. Their process hinges on a unique combination of the properties of organic semiconductors and the science of crystal nucleation and growth. First, pure crystals of so-called conjugated molecules — organic molecules that have alternate single and double covalent bonds — are known to be pretty good semiconductors^{3,4}.

Second, because these molecules are bound into a crystal through weak van der Waals forces, rather than through stronger covalent bonds, the energy needed to form crystals from them is moderate. Such crystals can therefore be formed at low temperatures, between room temperature and a few hundred degrees Celsius, that are compatible with the processing of large-area substrates.

Finally, thermodynamics teaches us that

many controllable parameters influence the energy barrier against the nucleation of crystals. These can be used to determine the location of the nucleation and growth of crystals on a surface. A known method for promoting local crystal nucleation is to 'stamp' a substrate with an extremely thin layer of molecules, known as a self-assembled monolayer, that is terminated by a chemical group with a particular affinity for the atoms or molecules of the desired crystal material⁵.

Briseno *et al.*¹ use a different method that is less dependent on chemical affinity. They start by stamping an organic layer of significant roughness onto the substrate. Roughness is known to increase surface energy, and the nucleation and growth of crystals can then be steered to occur selectively on the rougher areas. The authors show that the method can be used to grow many different semiconducting conjugated molecular crystals. These include molecules that are notoriously difficult to grow as films on substrates, such as rubrene⁶, and some with high vapour pressure, such as anthracene, that evaporate spontaneously in the high-vacuum environments in which thin-film deposition processes are generally conducted. But most importantly for device applications, the authors grow organic semiconductor crystals directly onto metal electrodes — forming a basic transistor — and show that these electrodes are functional after the process.

Considerable work is still needed to turn this technique into a truly viable semiconductor technology. In particular, the control of

crystal size and orientation will require further attention, as will the reproducibility and control of the quality of the electrical contacts between the semiconductor and the electrodes. The single-crystal technique might, uniquely, allow devices that are difficult to realize using thin polycrystalline or amorphous films of the same materials to be integrated onto foil. For example, most organic semiconductors are intrinsically ambipolar⁷, meaning that they allow the movement of both electrons and so-called holes — positive charges equivalent to the absence of an electron. This property has been difficult to exploit with thin organic semiconductor films, and single crystals might open new avenues towards this end.

So far, the family of technologies for use in electronics on flexible foil has numbered

thin-film organic semiconductors, thin-film metal-oxide semiconductors^{8,9}, and the twin silicons, amorphous and polycrystalline, among its members. Bao and colleagues¹ have just presented us with a new baby.

Paul Heremans is at IMEC, Kapeldreef 75, B3001 Leuven, Belgium.

e-mail: heremans@imec.be

1. Briseno, A. L. *et al.* *Nature* **444**, 913–917 (2006).
2. Huitema, H. E. A. *et al.* *Nature* **414**, 599 (2001).
3. de Boer, R. W. I. *et al.* *Phys. Status Solidi A* **201**, 1302–1331 (2004).
4. Pope, M. & Swenberg, C. E. *Electronic Processes in Organic Crystals and Polymers* (Oxford Univ. Press, 1999).
5. Aizenberg, J. *et al.* *Nature* **398**, 495–498 (1999).
6. Käfer, D. & Witte, G. *Phys. Chem. Chem. Phys.* **7**, 2850–2853 (2005).
7. Chua, L.-L. *et al.* *Nature* **434**, 194–199 (2005).
8. Nomura, K. *et al.* *Nature* **432**, 488–492 (2004).
9. Wang, L. *et al.* *Nature Mater.* **5**, 893–900 (2006).

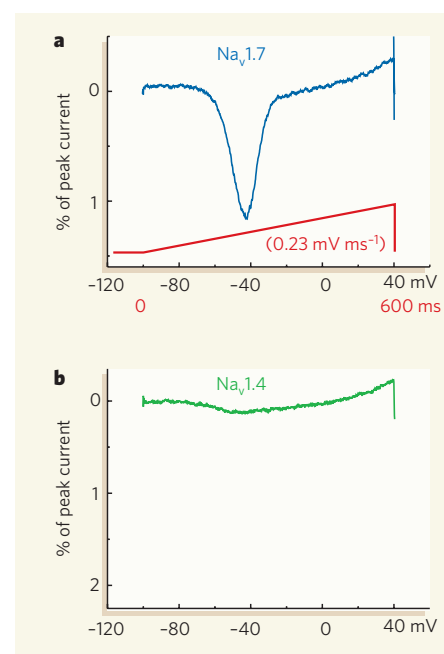


Figure 1 | Signalling through sodium channels. The graphs show changes in the potential difference across the cell membrane with gradually increasing stimulus amplitude. (The y-axis shows current as a percentage of the maximal response to an abrupt, large stimulus.) **a**, The $\text{Na}_v1.7$ sodium channel is found in the pain-sensing ends of nociceptive neurons. They produce a depolarization in the membrane (blue) in response to small, slow stimuli such as the ramp stimulus (red), causing the neuron to fire when the potential difference reaches a certain threshold. These channels therefore amplify the 'generator potentials' that are evoked by sensory stimuli and they set the gain on nociceptors. **b**, Other sodium channels, such as the $\text{Na}_v1.4$ sodium channel in muscle, do not produce these ramp responses. (Modified from ref. 3.)

NEUROBIOLOGY

A channel sets the gain on pain

Stephen G. Waxman

Nerve impulses that convey pain signals to the brain are produced by sodium channels in the neuronal membrane. Studies on people who are unable to feel pain identify one specific sodium channel as essential to the process.

"God...shouts in our pain. It is his megaphone..."

— C. S. Lewis

Pain can be useful — when it warns us that a hot object is burning us, for instance. Yet the amelioration of pain is essential to modern medicine, permitting surgery to take place and making many illnesses more bearable. Even so, some pain doesn't respond to current treatments; in particular neuropathic pain, which can occur in the absence of noxious stimuli following injury to the nervous system, and some types of inflammatory pain. In this issue, Cox *et al.* (page 894)¹ describe a rare inherited mutation that renders the people carrying it unable to feel any pain. The mechanism behind this deficit could aid the search for novel painkillers.

Where does pain come from, and how can we understand it so that we can tame it? Painful stimuli are conveyed in the form of trains of electrical impulses. These impulses travel from nociceptive (pain-signalling) dorsal root ganglion (DRG) neurons originating in the body's periphery, through ascending spinal pathways, to the brain. Ion channels in the cell membranes of these nociceptive neurons, including several different types of sodium channel, collaborate to produce such nerve impulses, although the relative roles of the different channels are not fully understood.

Ten different genes encode ten versions (isoforms) of the sodium-channel protein that all share a common structure but have different constituent amino-acid sequences.

One of these genes, *SCN9A*, encodes a sodium channel known as $\text{Na}_v1.7$, which is preferentially expressed at high levels in two types of neuron: nociceptive DRG neurons and sympathetic ganglion neurons, which are part of the involuntary, or 'autonomic', nervous system². $\text{Na}_v1.7$ is deployed at the endings of pain-sensing nerves (the nociceptors), close to areas where the impulse is initiated². Stimulation of the nociceptor nerve endings produces 'generator potentials' — small changes in the voltage across the neuronal membranes. The $\text{Na}_v1.7$ channel amplifies these membrane depolarizations³ (Fig. 1), and when the membrane potential difference reaches a certain threshold, the neuron fires.

Clues that $\text{Na}_v1.7$ is involved in pain came from the observation⁴ that DRG neurons in animal models of inflammatory pain show increased expression of $\text{Na}_v1.7$. Also, mice genetically engineered to lack $\text{Na}_v1.7$ specifically in their nociceptors show markedly reduced responses to inflammatory pain⁵. The painful inherited human neuropathy known as erythromelalgia, in which sufferers experience a severe burning pain in response to mild warmth, is due to mutations^{6–8} in $\text{Na}_v1.7$ that cause excessive channel activity^{9,10}. This suggests that $\text{Na}_v1.7$ sets the gain on pain signalling in humans.

Cox *et al.*¹ have now discovered *SCN9A* mutations that cause a loss of $\text{Na}_v1.7$ function in three families from Pakistan. Their observations link loss of $\text{Na}_v1.7$ function with a congenital inability to experience pain, adding to the

evidence indicting this sodium channel as an essential participant in human nociception. The profound insensitivity to pain in members of these families is remarkable because nociceptors contain many sodium-channel isoforms, so the remaining isoforms might have been expected to support at least some degree of nociception, even though $\text{Na}_v1.7$ is out of action.

The $\text{Na}_v1.7$ channel is present in some non-nociceptive DRG neurons¹¹, as well as in nociceptors. However, the people described by Cox *et al.* have no apparent deficits in non-nociceptive sensory functions, such as the ability to perceive touch, warmth, cold, tickle, pressure, or the position of their limbs (proprioception). This raises the question of whether $\text{Na}_v1.7$ channels have different roles in nociceptive and non-nociceptive DRG neurons. The pain insensitivity in these people seems to be more severe than the deficits seen in the mice lacking $\text{Na}_v1.7$ in their nociceptors⁵. So perhaps there are changes in upstream nociceptive centres in the spinal cord or brain of these people that contribute to their insensitivity to pain.

$\text{Na}_v1.7$ is present not only in DRG neurons, but also in sympathetic ganglion neurons². But

although total deletion of $\text{Na}_v1.7$ in mice results in perinatal death⁵, the people lacking functional $\text{Na}_v1.7$ showed no signs of dysfunction in their sympathetic ganglion neurons¹. Some mutations of $\text{Na}_v1.7$ seen in erythromelalgia depolarize both nociceptors and sympathetic ganglion neurons. Consistent with different functional roles in the two cell types, these mutations produce hypoexcitability in sympathetic ganglion neurons (because membrane depolarization inactivates all the sodium channels) and hyperexcitability in nociceptors (due to the selective presence of sodium channels that are relatively resistant to inactivation once depolarized)¹². Whether the absence of autonomic dysfunction in the families described by Cox *et al.*¹ is due to different functional roles of $\text{Na}_v1.7$ in sympathetic and nociceptive neurons, or to redundancy of sodium-channel subtypes in sympathetic neurons, is not known. It might also be due to increased sensitivity of upstream neurons induced by sympathetic hypoactivity, or to compensatory increases in the expression of other sodium-channel isoforms in the autonomic neurons (as can occur in other neuronal types in which mutations result in loss of other channels¹³).

The association of pain insensitivity with loss of function of a specific type of sodium channel might have therapeutic implications. Because $\text{Na}_v1.7$ is not present in cardiac muscle or neurons in the central nervous system, subtype-specific blockers of $\text{Na}_v1.7$ should not, in principle, have direct actions on these cells and so should have less-severe side-effects than current pain medications. Although $\text{Na}_v1.7$ channels are present in sympathetic neurons as well as nociceptors, the observations in the people studied by Cox *et al.* suggest that it may be possible to ameliorate pain by blocking $\text{Na}_v1.7$ without producing autonomic side-effects.

Whether $\text{Na}_v1.7$ blockers can abolish acute nociceptive signalling, thereby providing protection from pain similar to that afforded by local anaesthetics, but with fewer side-effects, remains to be determined. Also unknown is whether $\text{Na}_v1.7$ blockers can ameliorate chronic neuropathic or inflammatory pain without loss of the protection provided by acute nociception. If either of these possibilities turns out to be the case, it might be possible to develop a new generation of drugs that can effectively mute the megaphone of pain. ■

Stephen G. Waxman is in the Department of Neurology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA and the Center for Neuroscience and Regeneration Research, VA Hospital, West Haven, Connecticut.
e-mail: stephen.waxman@yale.edu

6. Drenth, J. P. *et al.* *Am. J. Hum. Genet.* **68**, 1277–1282 (2001).
7. Yang, Y. *et al.* *J. Med. Genet.* **41**, 171–174 (2004).
8. Waxman, S. G. & Dib-Hajj, S. D. *Trends Mol. Med.* **11**, 555–562 (2005).
9. Cummins, T. R., Dib-Hajj, S. D. & Waxman, S. G. *J. Neurosci.* **24**, 8232–8236 (2004).

10. Dib-Hajj, S. D. *et al.* *Brain* **128**, 1847–1854 (2005).
11. Djouhri, L. *et al.* *J. Physiol. (Lond.)* **546**, 565–576 (2003).
12. Rush, A. M. *et al.* *Proc. Natl Acad. Sci. USA* **103**, 8245–8250 (2006).
13. Van Wart, A. & Matthews, G. *J. Neurosci.* **26**, 7172–7180 (2006).

CONDENSED-MATTER PHYSICS

Up the magnetic pressure

Shaun Fisher and George Pickett

Observations of a phenomenon known as the magnetic fountain effect in superfluid helium are not just beautiful experiments, but could also supply a tool for studying many other exotic magnetic phenomena.

At temperatures just a few millikelvin above absolute zero, the light helium isotope ^3He , in its liquid state, develops exotic superfluid properties¹. Uniquely, this superfluidity allows the frictionless transport of not only mass, but also magnetism. On page 909 of this issue², Yamaguchi *et al.* exploit the interplay between these two properties to produce a device based on superfluid ^3He that induces magnetism from applied pressure, and vice versa.

Superfluidity — flow without friction — occurs when a macroscopic number of particles are made to occupy the lowest available energy state, or ground state, of a system. Once they form such a ‘condensate’, the particles behave collectively as a single quantum-mechanical entity, resulting in uninhibited flow. The exact make-up of the condensate is determined by the quantum-mechanical division of particles into two categories according to the angular momentum of their intrinsic spin (‘spin’ for short). Bosons have spin values that are integer multiples of an elementary quantum (known as the reduced Planck constant and written $\hbar$). Fermions, on the other hand, have half-integer spin values; the electrons, protons and neutrons that make up atoms all fall into this latter category.

The simplest superfluids are formed by bosons through the process known as Bose–Einstein condensation. A common example is superfluid ^4He , created at temperatures below around 2 kelvin (ref. 3). A ^4He atom consists of two half-integer-spin electrons orbiting a nucleus of four half-integer-spin nucleons — two protons and two neutrons. These spins combine to give a total atomic spin of zero, making the ^4He atom a boson. More ideal Bose–Einstein condensates can be formed from ultracold dilute gases⁴.

In contrast to ^4He , the odd number of nucleons in the nucleus of ^3He , which comprises two protons and a neutron, gives the ^3He atom as a whole a half-integer spin, making it a fermion. Unlike bosons, fermions obey the rule known as the Pauli exclusion principle, which allows no more than one particle to occupy any one quantum state. That would seem to forbid

the formation of a superfluid condensate from fermions. At low temperatures, however, fermions may pair up to make integer-spin bosons from which a condensate can form. This process of pair condensation occurs among conduction electrons in a solid, where it leads to the phenomenon of superconductivity, and it has recently been shown to occur in some ultracold atomic gases⁵.

In liquid ^3He , the superfluid condensation that occurs below a few millikelvin leads to some exotic behaviour. The condensate pairs have non-zero spin and orbital angular momenta that point in particular directions. This second attribute imparts directionality to the fluid, as the fluid responds differently to disturbances that are parallel or perpendicular to the directions of the angular momenta. These directions may vary in different regions of the fluid, producing momentum ‘textures’. Because the pair condensate carries not only mass, but also spin and orbital angular momentum, superfluidity in ^3He can, under certain conditions, be regarded as the frictionless flow of all three of these material properties⁶. As spin is intimately related to the magnetic response of a particle, superfluid ^3He thus also supports the frictionless transport of magnetization.

Yamaguchi *et al.*² exploit these qualities of ^3He to demonstrate for the first time the magnetic version of an exotic thermo-mechanical phenomenon seen in other superfluids. This effect is commonly known as the fountain effect, and occurs at temperatures above absolute zero when some fraction of the atoms is thermally excited into higher energy levels. As these atoms are not part of the condensate, they form an ordinary, classical fluid component that has non-zero viscosity (viscosity being essentially a measure of the frictional resistance that a fluid engenders).

Suppose now that this mix of classical fluid and superfluid is pushed through a fine tube or porous plug by applying a pressure gradient. The zero-viscosity superfluid component will flow freely, whereas the normal-fluid component is hindered by its viscosity. This plug is known as a superleak, and acts to filter out the

1. Cox, J. J. *et al.* *Nature* **444**, 894–898 (2006).
2. Toledo-Aral, J. J. *et al.* *Proc. Natl Acad. Sci. USA* **94**, 1527–1532 (1997).
3. Cummins, T. R., Howe, J. R. & Waxman, S. G. *J. Neurosci.* **18**, 9607–9619 (1998).
4. Black, J. A. *et al.* *Pain* **108**, 237–247 (2004).
5. Nassar, M. A. *et al.* *Proc. Natl Acad. Sci. USA* **101**, 12706–12711 (2004).

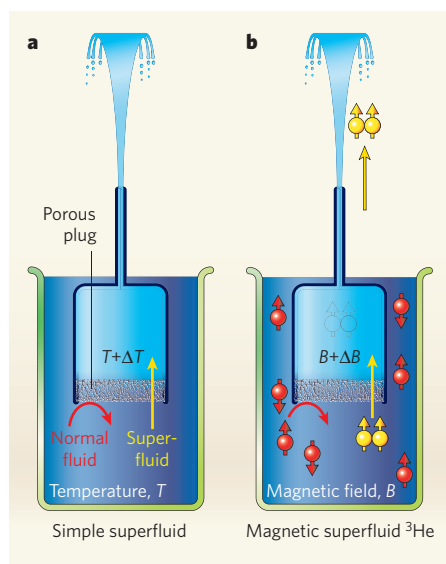


Figure 1 | The magnetic fountain effect. **a**, A simple bosonic superfluid (^4He , for example) comprises a condensed superfluid component mixed into the normal fluid state. When a temperature gradient ΔT is applied across a flow restriction such as a porous plug (a 'superleak'), the superfluid component flows up through the restriction to equalize the temperature, whereas the normal-fluid component is prevented from passing by its viscosity. The result is a 'fountain' of superfluid. **b**, In ^3He , the superfluid component consists of pairs of atoms (yellow), because the fermionic nature of ^3He atoms prevents them from condensing singly into a superfluid state. In the A_1 phase of ^3He studied by Yamaguchi *et al.*², the superfluid pairs are highly magnetized, with their spins aligned with a high magnetic field, whereas the unpaired atoms (red) are not magnetized. In this case, the magnetized spin pairs will move to equalize a magnetization difference generated by a magnetic field gradient ΔB , thus supplying the driving force for a magnetic fountain.

normal-fluid component. Because the superfluid condensate carries no thermal energy, the fluid downstream of the plug cools, whereas the fluid that has yet to pass through the plug warms up as the relative concentration of thermally excited normal fluid increases. This mechano-thermal effect, in which an applied pressure gradient results in a temperature gradient over the plug, is well known in superfluid ^4He . The fountain effect is the reverse effect, and occurs when a temperature gradient is applied across the superleak. This induces a flow as the superfluid moves to equalize the temperatures, and thus a pressure difference that, when directed upwards, produces a fountain³ (Fig. 1a).

Yamaguchi and colleagues' magnetic analogy of the fountain effect involves the so-called A_1 phase of superfluid ^3He , which exists only in high magnetic fields. In the A_1 phase, all the atom pairs of the superfluid condensate align their spins along the field, whereas the unpaired atoms do not. The superleak allows only the highly magnetized superfluid component to pass, and thus acts as a spin filter: a mechanical

flow produced by a pressure difference across the superleak will cause the fluid downstream to become more magnetized. Conversely, a difference in magnetization, generated by a field gradient, will induce a pressure difference across the superleak, because the superfluid flows to equalize magnetization (Fig. 1b).

The experiments² require sensitive measurements in very high magnetic fields and at extremely low temperatures. Having observed the effect, the authors went on to study magnetic relaxation processes using the device: a loss of magnetization will be reflected in a loss of fountain pressure, measured by the authors through the deflection of a diaphragm. Intriguingly, their findings suggest that the superfluid A_1 phase in liquid helium is not perfectly magnetized, as had been assumed. Rather, a small fraction of the condensate pairs seems to be orientated against the direction of the field. Consequently, the simplified concepts commonly used to characterize the A_1 phase might have to be revised.

Although the magnetic fountain effect by itself is a beautiful demonstration of a mechanical spin filter, it could also prove to be a useful tool for studying a wide range of magnetic

phenomena, such as the relaxation of magnetization in high magnetic fields. There is also great potential for future technological applications based on this system or analogous electronic systems. Some exotic superconductors have structures similar to that of superfluid ^3He and might also exhibit a magnetic fountain effect. If such materials can be integrated into devices, then frictionless devices for manipulating spin that rely on externally applied quantities such as magnetic field can readily be envisaged. Such properties could well provide access to a new range of devices and micromachines in the rapidly growing field of spin-based electronics. ■

Shaun Fisher and George Pickett are in the Department of Physics, Lancaster University, Lancaster LA1 4YB, UK.

1. Volovik, G. E. *Exotic Properties of Superfluid ^3He* (World Scientific, Singapore, 1992).
2. Yamaguchi, A., Kobayashi, S., Ishimoto, H. & Kojima, H. *Nature* **444**, 909–912 (2006).
3. Guénault, A. M. *Basic Superfluids* (Taylor & Francis, London, 2003).
4. Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. *Science* **269**, 198–201 (1995).
5. Zwierlein, M. W., Schunck, C. H., Schirotzek, A. & Ketterle, W. *Nature* **442**, 54–58 (2006).
6. Fisher, S. N. & Suramlishvili, N. J. *Low Temp. Phys.* **141**, 111–141 (2005).

DIABETES

Kicking off the insulin cascade

Catherine Jackson

Inhibition of the insulin-signalling pathway leads to insulin resistance, an early step in the development of type 2 diabetes. A novel family of protein activators seems to act near the pathway's inception.

Many of the cellular processes that are vital for human health are regulated by the insulin/insulin-like growth factor signalling pathway, including glucose homeostasis, fat metabolism, cell growth and differentiation, and ageing^{1,2}. Defects in this pathway are central to the development of obesity and related diseases such as diabetes, in cancer and in the process of ageing. In this issue, Hafner *et al.* (page 941)³ and Fuss *et al.* (page 945)⁴ report the exciting discovery of a group of factors involved in the insulin-signalling pathway — the ARNO/cytohesin family of Arf activators.

The Arf proteins are molecular switches that control the assembly of multiprotein complexes onto membranes to regulate protein trafficking throughout the cell and membrane-associated signalling events⁵. Arf proteins are inactive when bound to guanine nucleotide diphosphate (GDP) and are active when bound to guanine nucleotide triphosphate (GTP). For their activation, Arf proteins rely on guanine nucleotide exchange factors (GEFs), which catalyse the replacement of GDP with GTP (ref. 5). The ARNO/cytohesin proteins are Arf GEFs that have a characteristic catalytic domain called a

Sec7 domain. They are part of a larger family of Sec7-domain proteins that have key roles in membrane trafficking and signal-transduction pathways throughout the cell^{5,6}. There are four ARNO/cytohesin family members in humans: cytohesin-1, ARNO/cytohesin-2, GRP1/cytohesin-3 and cytohesin-4 (ref. 6).

There is only one known membrane-permeable, small-molecule inhibitor of the Arf GEFs, the fungal toxin brefeldin A, which is highly specific and targets only three of the Arf GEFs (refs 5, 7). The ARNO/cytohesin proteins, however, have only a very weak affinity for brefeldin A, and their activity is not inhibited by this drug^{6,7}.

Previously, Michael Famulok and colleagues reported⁸ the discovery of an RNA aptamer — a small RNA molecule that binds specifically to a particular protein — that inhibited the guanine nucleotide exchange activity of the ARNO/cytohesin proteins, but not the three GEFs targeted by brefeldin A. Now the same group (Hafner *et al.*³) reports an innovative method for 'converting' an RNA aptamer into a membrane-permeable molecule, and the consequent discovery of the first small-molecule inhibitor

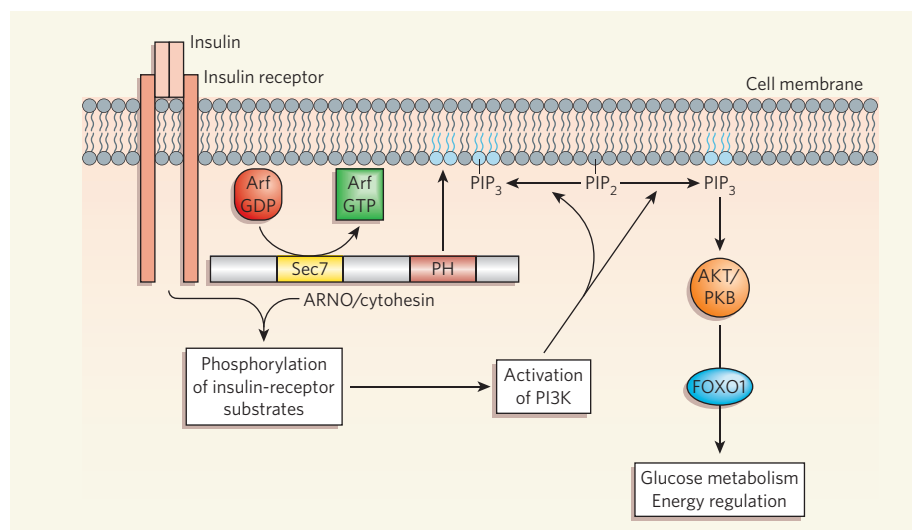


Figure 1 | The insulin-signalling cascade. The binding of insulin to its receptor on the cell surface leads to a cascade of events that eventually activates the FOXO1 gene regulatory factor, which controls the expression of genes involved in glucose metabolism and energy regulation. Hafner *et al.*³ and Fuss *et al.*⁴ show that the ARNO/cytohesin family of proteins is involved in the pathway, possibly near its start. These proteins are guanine nucleotide exchange factors, using their Sec7 domain to catalyse the exchange of GDP for GTP to activate the Arf proteins that mediate the formation of protein complexes on cellular membranes. One of the steps in insulin signalling is the conversion of the lipid molecule PIP₂ to PIP₃, a lipid that is specifically recognized by protein domains called pleckstrin homology (PH) domains.

specific for the ARNO/cytohesin family of Arf GEFs. Their approach involves attaching a fluorescent tag to the RNA aptamer, then screening a large library of small molecules to find those that could displace the fluorescent RNA molecule from its protein target. The authors went on to show that the candidate molecule identified by this screen, SecinH3, does indeed bind to the ARNO and cytohesin proteins with high affinity and inhibits their exchange activity on Arf. By contrast, it binds much less efficiently to other Sec7-domain proteins, and fails to inhibit their exchange activity.

Meanwhile, Fuss *et al.*⁴ were studying fruitflies that carry mutations in the single gene (called *steppke*) encoding ARNO/cytohesin in this organism, and they discovered that the mutant flies were smaller and developed more slowly than their normal counterparts⁴. These characteristics can arise from defects in insulin signalling, and indeed, the authors found a profile of gene regulation in these mutant flies consistent with a defect in the insulin cascade. Moreover, when they treated the flies with SecinH3, they found nearly identical traits to those caused by mutations in the *steppke* gene.

These results prompted Famulok's group to test the effect of SecinH3 on insulin signalling in mammalian cells. Insulin and insulin-related growth factor bind to a family of receptors. These receptors are protein kinases; that is, they are able to attach phosphate chemical groups to target proteins — a common means of passing information from one protein to the next along a cell-signalling pathway. And so begins the complex cascade of insulin signalling² (Fig. 1). Binding of insulin to its receptor stimulates phosphorylation of a set of 'insulin receptor substrate' proteins, which in turn

recruit and activate a lipid kinase, phosphatidylinositol 3-kinase (PI3K). PI3K converts the phosphoinositide PIP₂ to PIP₃, a lipid species that is specifically recognized by proteins with pleckstrin-homology (PH) domains, notably the protein kinase AKT/PKB (ref. 2). AKT/PKB phosphorylates a number of targets including FOXO1, a factor that regulates the expression of insulin-sensitive genes.

In cultured liver cells (HepG2), treatment with SecinH3 produced a gene-expression profile remarkably similar to that of cells with defects in the insulin receptor. Moreover, mice fed SecinH3 had traits similar to mice genetically engineered to have no insulin receptor in the liver tissue³. In particular, the SecinH3-fed mice showed signs of insulin resistance, one of the first steps in the development of type 2 diabetes and a cause of 'metabolic syndrome', which is a growing global health problem.

Where in the insulin-signalling pathway does ARNO/cytohesin act? In fruitflies, Fuss *et al.*⁴ show genetically that the cytohesin encoded by *steppke* acts upstream of PI3K. In mammalian cells, Hafner *et al.*³ found that SecinH3 inhibited phosphorylation of insulin receptor substrate 1 by the insulin receptor. Furthermore, co-immunoprecipitation experiments showed that in cells the insulin receptor is physically associated with ARNO and cytohesin-3, as well as one of their substrates, Arf6, whose activation is inhibited by SecinH3. These results indicate that the ARNO/cytohesin proteins act very early in the insulin cascade.

The ARNO/cytohesin proteins all have a very similar modular structure, with an amino-terminal coiled-coil domain, the central Sec7 domain, and a PH domain at the carboxy terminus that specifically binds to PIP₂ and/or PIP₃

(Fig. 1)⁶. Remarkably, variants of the cytohesins in mammalian cells that differ by only a single glycine residue in their PH domains have different capacities to bind PIP₂ versus PIP₃ (ref. 9). In fruitflies, all cytohesin variants seem to have the PH domain that binds with much greater affinity to PIP₃ than to PIP₂. The ARNO/cytohesin proteins, as well as Arf1 and Arf6, are recruited to the cell membrane by the insulin-stimulated production of PIP₃ (refs 9–11). Cells have no PIP₃ before stimulation with insulin or other growth factors. So, this raises the interesting possibility that production of PIP₃ by the insulin-signalling cascade creates a positive-feedback loop that leads to recruitment of additional ARNO/cytohesin molecules to the PIP₃-enriched regions of the membrane.

Metabolic syndrome develops as a consequence of obesity and insulin resistance, and has grown to epidemic proportions in recent years with the rapid escalation in obesity throughout the world¹. The discovery of a role for ARNO/cytohesin proteins in insulin signalling provides a new target for the development of drugs to combat this growing health problem. The method developed by Famulok and colleagues provides a powerful way of identifying small-molecule inhibitors of therapeutic value. An aptamer library has much more variability than a standard drug library, so it can be used in an initial screening with a much higher probability of success. The method presented can then be used to easily and rapidly screen multiple small-molecule libraries to find drug-like compounds with similar inhibitory potential to that of the original aptamer.

The results presented by Hafner, Fuss and colleagues^{3,4} open up a fresh area for investigation, with many questions to be answered. What are the Arf substrates that ARNO/cytohesin GEFs act on in the insulin-signalling pathway? How does ARNO/cytohesin regulate the protein kinase activity of the insulin receptor on its substrate proteins? Are defects in ARNO/cytohesin proteins the basis for insulin resistance in type 2 diabetes? Given how important this signalling pathway is to human health, we can look forward to answers appearing quickly. ■

Catherine Jackson is in the Laboratory of Enzymology and Structural Biochemistry, Bat 34, CNRS, 91198 Gif sur Yvette, France. e-mail: catherine.jackson@lebs.cnrs-gif.fr

1. Biddinger, S. B. & Kahn, C. R. *Annu. Rev. Physiol.* **68**, 123–158 (2006).
2. Taniguchi, C. M., Emanuelli, B. & Kahn, C. R. *Nature Rev. Mol. Cell Biol.* **7**, 85–96 (2006).
3. Hafner, M. *et al. Nature* **444**, 941–944 (2006).
4. Fuss, B., Becker, T., Zinke, I. & Hoch, M. *Nature* **444**, 945–948 (2006).
5. Jackson, C. L. & Casanova, J. E. *Trends Cell Biol.* **10**, 60–67 (2000).
6. Cox, R., Mason-Gamer, R. J., Jackson, C. L. & Segev, N. *Mol. Biol. Cell* **15**, 1487–1505 (2004).
7. Zeesh, J. C. *et al. J. Biol. Chem.* **281**, 11805–11814 (2006).
8. Mayer, G. *et al. Proc. Natl Acad. Sci. USA* **98**, 4961–4965 (2001).
9. Cullen, P. J. & Chardin, P. *Curr. Biol.* **10**, R876–R878 (2000).
10. Czech, M. P. *Cell* **100**, 603–606 (2000).
11. Li, H. S. *et al. BMC Cell Biol.* **4**, 13 (2003).

BRIEF COMMUNICATIONS

A microworld in Triassic amber

Amber as old as the first dinosaurs captured the diversity of microbial life 220 million years ago.

Amber provides an effective medium for conservation of soft-bodied microorganisms^{1–6}, but finds older than 135 million years are very rare and have not so far contained any microbial inclusions. Here we describe 220-million-year-old droplets of amber containing bacteria, fungi, algae and protozoans that are assignable to extant genera. These inclusions provide insight into the evolution and palaeoecology of Lower Mesozoic microorganisms: it seems that the basal levels of food webs of terrestrial communities (biocoenoses) have undergone little or no morphological change from the Triassic to the Recent.

The largest known deposit of Triassic amber was discovered near the town of Cortina d'Ampezzo in the Italian Dolomites (southern Alps)⁷. Several thousand millimetre-sized drop-shaped pieces of amber (Fig. 1a) per square metre were preserved in a palaeosol (for details, see supplementary information).

Bacteria were the most abundant microbes in these drops. Anamorphic fungi including some conidia-producing fungi that might be related to the extant genus *Ramularia* were present. Besides tiny chlorophytes and resting stages of several algae, the oldest known fossils of vegetative cells of conjugatophytes were also evident inside the droplets (Fig. 1b). These desmids are similar to the extant genus *Cosmarium* and were embedded at different stages of asexual reproduction. Ciliate protozoans are well preserved and, *inter alia*, a ciliate of the modern genus *Coleps* (family Colepidae) is identifiable (Fig. 1c).

Testate amoebae are represented by several species of the families Centropyxidae and Difflugiidae, and this is the earliest evidence of non-marine testaceans. Judging by its shape, size and the eccentric round pseudostome, the inclusion shown in Fig. 1d is morphologically identical to extant representatives of *Centropyxis hirsuta* Deflandre and therefore represents the earliest occurrence of a species of protozoan that survives today. As in extant individuals that live in mosses, soil or on bark, no spines are evident on the fossil shell⁸.

Representatives of all trophic levels of the microbiocoenosis were preserved in these samples of Triassic amber: bacteria, and phototrophic algae as producers, protozoans as consumers, and fungi as decomposers. Bacteria are the food source of testate amoebae and some

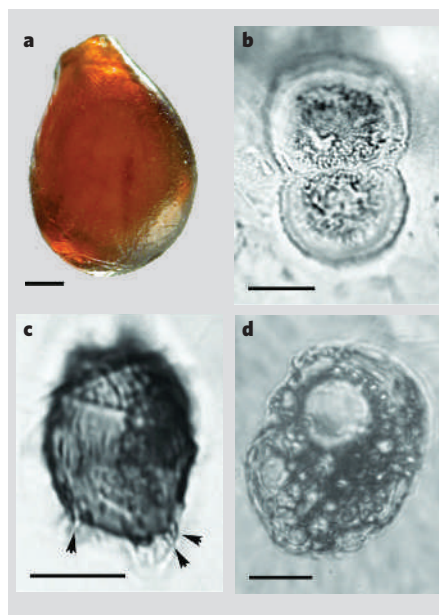


Figure 1 | Triassic amber with microinclusions, from the Dolomites in northern Italy. a, Drop-shaped piece of amber with a broken tip. Scale bar, 500 μm . **b**, Alga similar to the extant desmid genus *Cosmarium*. The flattened cell consists of two almost symmetrical half-cells, each of which contains a large disc-shaped chloroplast (dark area). Scale bar, 10 μm . **c**, Ciliate of the genus *Coleps*. The barrel-shaped cell body shows rows of largely reniform perforations at its surface. Arrows indicate the typical posterior spines. Scale bar, 5 μm . **d**, Testate amoeba *Centropyxis hirsuta*. Scale bar, 10 μm . (For further details, see supplementary information.)

ciliates; microalgae may be grazed upon by ciliates as well. Extant species of the ciliate genus *Coleps* also prey on other ciliates, flagellates, and on small metazoans. Freshwater rhizopods occur in high densities in extant aquatic and terrestrial habitats and participate to a high degree in secondary production. Our discovery of associations of several species in just one piece of amber indicates that these representatives of different testacean families had settled in such habitats, outside large bodies of water, by the time of the Triassic.

Humid coastal forests of the archaic conifer family Cheirolepidiaceae at the northern rim of the Tethys Sea are likely to be the amber source⁹. The drop-shaped pieces of amber solidified on their resin-bearing plants and later fell onto

the soil. The microbes must therefore have been living on plants; no accompanying soil organisms were enclosed. The occurrence of conjugatophytes and aquatic ciliates indicates that they probably lived in limnetic–terrestrial habitats, such as wet bark or small water-filled treeholes (phytotelmata).

Microorganisms of limnetic and terrestrial biocoenoses are important in modern ecosystems, but until now their fossil record has been poor. Our findings show that different genera, and even species, of microbial taxa have been able to survive geological epochs. Higher levels in food webs, on the other hand, have been shaped by environmental changes, such as those that caused the mass extinction at the Cretaceous–Palaeogene boundary. Unchanged since the Lower Mesozoic, protozoans survived the entire era of the dinosaurs, as well as the diversification of angiosperms, birds and mammals.

Alexander R. Schmidt*, **Eugenio Ragazzi†**, **Olimpia Coppellotti‡**, **Guido Roghi§**

*Museum für Naturkunde der Humboldt-Universität zu Berlin, 10115 Berlin, Germany
e-mail: alexander.schmidt@museum.hu-berlin.de

†Department of Pharmacology and Anaesthesiology, University of Padova, 35131 Padova, Italy

‡Department of Biology, University of Padova, 35131 Padova, Italy

§Institute of Geosciences and Earth Resources, CNR and Department of Geology, Paleontology and Geophysics, University of Padova, 35137 Padova, Italy

- Waggoner, B. M. *Lethaia* **27**, 77–84 (1994).
- Schönborn, W., Dörfelt, H., Foissner, W., Krienitz, L. & Schäfer, U. *J. Eukaryot. Microbiol.* **46**, 571–584 (1999).
- Dörfelt, H. & Schäfer, U. *Hoppea* **61**, 785–793 (2000).
- Rikkinen, J. & Poinar, G. O. Jr *Mycol. Res.* **104**, 7–15 (2000).
- Schmidt, A. R., Schönborn, W. & Schäfer, U. *Palaeontology* **47**, 185–197 (2004).
- Dörfelt, H. & Schmidt, A. R. *Palaeontology* **49**, 1013–1017 (2006).
- Roghi, G., Ragazzi, E. & Gianolla, P. *Palaios* **21**, 143–154 (2006).
- Schönborn, W. *Limnologia* **1**, 111–182 (1962).
- Roghi, G. *Rev. Palaeobot. Palynol.* **132**, 1–35 (2004).

Supplementary information accompanies this communication on Nature's website.

Received 18 August; accepted 4 October 2006.

Competing financial interests: declared none.
doi:10.1038/444835a

MOLECULAR EPIDEMIOLOGY

HIV-1 and HCV sequences from Libyan outbreak

In 1998, outbreaks of human immunodeficiency virus type 1 (HIV-1) and hepatitis C virus (HCV) infection were reported in children attending Al-Fateh Hospital in Benghazi, Libya. Here we use molecular phylogenetic techniques to analyse new virus sequences from these outbreaks. We find that the HIV-1 and HCV strains were already circulating and prevalent in this hospital and its environs before the arrival in March 1998 of the foreign medical staff (five Bulgarian nurses and a Palestinian doctor) who stand accused of transmitting the HIV strain to the children.

Almost half of the 111 children studied in the early months after the discovery of the outbreak showed evidence of both HIV-1 and HCV infection¹. Of 418 children eventually affected by these viruses, 248 were referred to European hospitals^{1,2}. Sequence analysis of 51 children classified the HIV-1 infection as the strain CRF02_AG; HCV infection was classified as genotype 4 or subtype 1a in 15 children^{1,2}.

We studied HIV-1 *gag* gene sequences from 44 affected children, plus 61 HCV *E1E2* gene sequences that span the HCV hypervariable region (for methods, see supplementary information). By using these data in an evolutionary analysis, we could place a real timescale on the transmission history of the outbreaks.

We collated all available reference strains that were closely related to the sequences from the Al-Fateh Hospital, then estimated and assessed phylogenies using algorithmic, bayesian and maximum-likelihood methods

(for details, see supplementary information). The HIV-1 sequences from the hospital form a well supported monophyletic cluster within the CRF02_AG clade, indicating that the outbreak arose from one CRF02_AG lineage. The cluster is closest to three west African reference sequences (Fig. 1a), the basal location of which suggests that the Al-Fateh Hospital lineage arrived in Libya from there. The branch length leading to the Al-Fateh Hospital cluster is perfectly typical; hence the Al-Fateh Hospital strain is not unusually divergent².

In an equivalent HCV phylogenetic analysis, the HCV sequences from the hospital formed three monophyletic clusters containing 11 subtype-4a sequences, phylogenetically placed among Egyptian subtype 4a lineages; 22 sequences most closely related to a Cameroonian genotype-4 strain; and 24 sequences belonging to the worldwide and prevalent subtype 1a; four remaining sequences belong to genotype 4 (Fig. 1b, c; see supplementary information).

Epidemiological linkage of the HIV-1 and HCV clusters from Al-Fateh Hospital with sequences from sub-Saharan Africa is to be expected, given the large number of migrants within or passing through Libya³; indeed, the Libyan authorities have expressed concern about the risk of introduction of HIV/AIDS and hepatitis as a result of this migration⁴. In addition, HCV genotype 4 is endemic to central Africa and the Middle East⁵⁻⁷, and subtype 4a is exceptionally prevalent in neighbouring Egypt^{8,9}.

Virus sequences also contain temporal information about the date of origin and age of epidemics¹⁰. We therefore comprehensively analysed the evolution of the Al-Fateh Hospital clusters using an established bayesian Markov chain Monte Carlo (MCMC) approach^{9,10} that appropriately accounts for estimation uncertainty. We estimated three parameter values for each cluster: the date of its most recent common ancestor; the probability that its most recent common ancestor was more recent than 1 March 1998; and the percentage of its lineages that already existed before 1 March 1998. (These values are conservative, because cluster origins could be older than the most recent common ancestor, but not younger.) To avoid model selection bias, we used a range of applicable models.

We found that, irrespective of which model was used, the estimated date of the most common recent ancestor for each cluster predated March 1998, sometimes by many years (Fig. 2).

In most analyses, the probability that the clusters from the Al-Fateh Hospital originated after that time was almost zero (for details, see supplementary information). For the three HCV clusters, the percentage of lineages already present before March 1998 was about 70%; the equivalent percentage for the HIV-1 cluster was estimated at about 40%.

Our results support the existing nosocomial transmission scenario^{1,11} and suggest that Al-Fateh Hospital had a long-standing infection-control problem. The earlier origin and

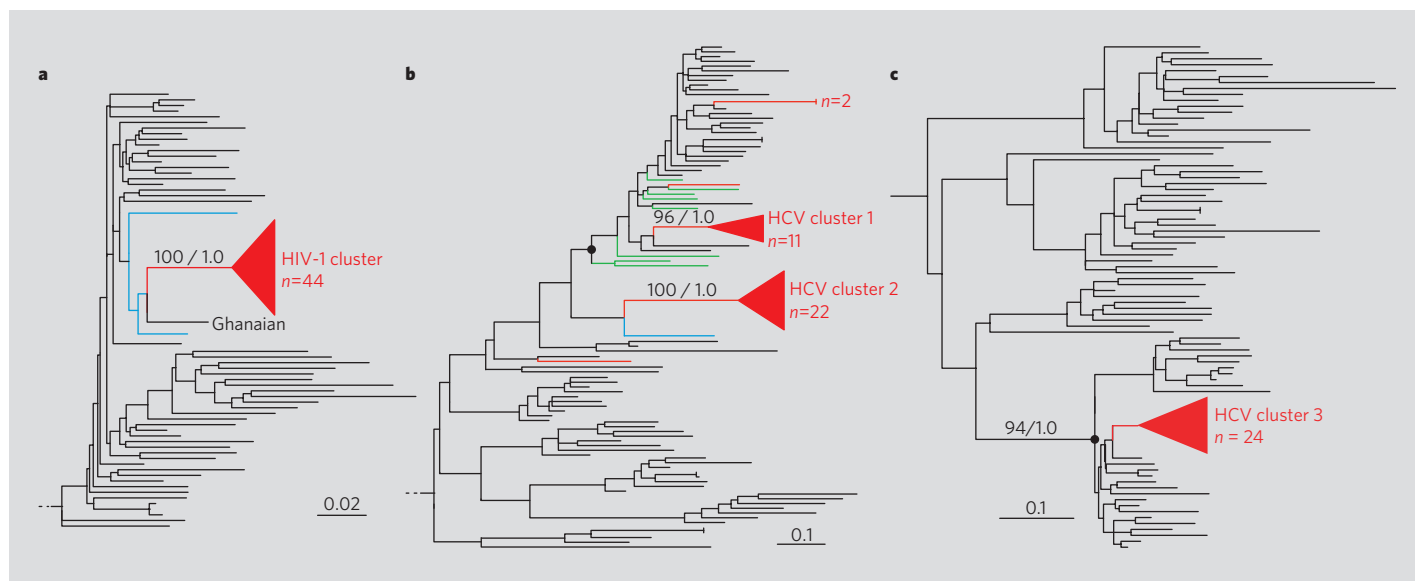


Figure 1 | HIV-1 and HCV sequences from 1998 Al-Fateh Hospital (AFH) outbreak. **a–c**, Estimated maximum-likelihood phylogenies for HIV-1 CRF02_AG (**a**), HCV genotype 4 (**b**) and HCV genotype 1a (**c**). Source of sequences used for analysis: AFH, red; Egypt, green; Cameroon, blue. Black circles mark the common ancestor of HCV subtype 4a and 1a; numbers above AFH lineages give clade support values using bootstrap and bayesian methods, respectively. Scale bar units are nucleotide substitutions per site. For visual clarity, AFH clusters are represented by triangles and some non-informative reference strains are excluded.

greater number of HCV clusters than HIV-1 clusters reflect the higher transmissibility of HCV compared with HIV-1 by such routes¹². Crucially, we have shown that the HIV-1 and HCV strains responsible were being spread and transmitted among individuals attending the hospital before March 1998, indicating that many of the transmissions giving rise to the infection clusters must have already occurred before the foreign medical staff arrived.

Tulio de Oliveira*, **Oliver G. Pybus***, **Andrew Rambaut†**, **Marco Salemi‡**, **Sharon Cassol§**, **Massimo Ciccozzi||**, **Giovanni Rezza||**, **Guido Castelli Gattinara¶**, **Roberta D'Arrigo#**, **Massimo Amicosante****, **Luc Perrin††**, **Vittorio Colizzi‡‡**, **Carlo Federico Perno§§**, **Benghazi Study Group|||**

*Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

e-mail: tulio.deoliveira@zoo.ox.ac.uk

†Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

‡College of Medicine, University of Florida, Gainesville, Florida 32610, USA

§Faculty of Health Sciences, University of Pretoria, 0001, Pretoria, South Africa

||Department of Infectious Diseases, Istituto Superiore di Sanità, 00161 Rome, Italy

¶Department of Immunology and Infectious Diseases, Bambino Gesù Children's Hospital, 00165 Rome, Italy

#Spallanzani Institute for Infectious Diseases, 00149 Rome, Italy

**Department of Internal Medicine, University of Rome "Tor Vergata", 00133 Rome, Italy

††Geneva University Hospital, 1211, Geneva, Switzerland

‡‡Department of Biology, University of Rome "Tor Vergata", 00133 Rome, Italy

§§Department of Experimental Medicine and Biochemical Sciences, University of Rome "Tor

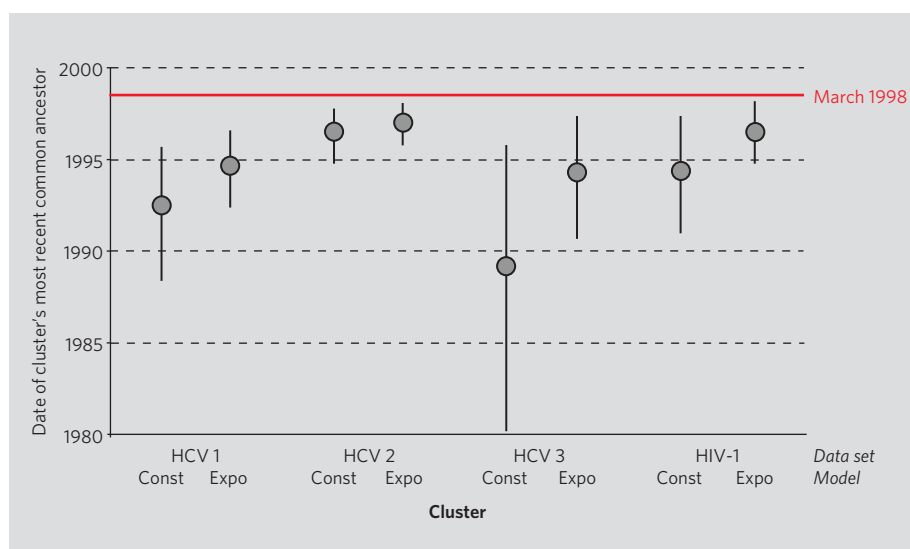


Figure 2 | Estimated dates of the most recent common ancestor for each cluster. Results obtained by using different evolutionary models. Vertical lines show the 95% highest posterior density intervals. Red line shows time of arrival of the foreign staff in March 1998. For further details, see supplementary information. 'Const', constant size; 'Expo', exponential growth.

Vergata", 00173 Rome, Italy

|||For full listing, see supplementary information.

1. Yerly, S. *et al. J. Infect. Dis.* **184**, 369–372 (2001).
2. Visco-Comandini, U. *et al. AIDS Res. Hum. Retrovir.* **18**, 727–732 (2002).
3. Human Rights Watch *Stemming the Flow* Vol. 18, No. 5 (2006); <http://www.hrw.org/reports/2006/libya0906>
4. European Commission *Technical Mission to Libya on Illegal Immigration*, 27 Nov–6 Dec 2004 (2006); <http://www.statewatch.org/news/2005/may/eu-report-libya-ill-imm.pdf>
5. Ndjomou, J., Pybus, O. G. & Matz, B. *J. Gen. Virol.* **84**, 2333–2341 (2003).
6. Smith, D. B. *et al. J. Gen. Virol.* **78**, 321–328 (1997).
7. Pybus, O. G. *et al. Science* **292**, 2323–2325 (2001).
8. Frank, C. *et al. Lancet* **355**, 887–891 (2000).
9. Pybus, O. G., Drummond, A. J., Nakano, T., Robertson, B. H.

& Rambaut, A. *Mol. Biol. Evol.* **20**, 381–387 (2003).

10. Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. *PLoS Biol.* **4**, e88 (2006).

11. Montagnier, L. & Colizzi, V. *Report on the Benghazi Outbreak* (2006); <http://www.nature.com/nature/journal/v443/n7114/extref/montagnier.pdf>

12. Goldmann, D. A. *J. Allergy Clin. Immunol.* **110** (suppl.), 21–26 (2002).

Supplementary information accompanies this communication on Nature's website.

Received 4 November; accepted 24 November 2006. Published online 6 December 2006.

Competing financial interests: declared none
doi:10.1038/nature444836a

BRIEF COMMUNICATIONS ARISING online
♦ www.nature.com/bca see Nature contents.

EVOLUTIONARY GENETICS

Evolution of mate choice in the wild

Arising from: A. Qvarnström, J. E. Brommer & L. Gustafsson *Nature* **441**, 84–86 (2006)

Qvarnström *et al.*¹ test whether the preference of female collared flycatchers (*Ficedula albicollis*) for males with large forehead patches could have evolved as a by-product of selection acting on male patch size². They find that the crucial genetic correlation between female choice and male patch size is not significant, and conclude that preference for large patches must have been shaped directly by selection. However, their use of the patch size of a female's social partner as a measure of choice is incomplete, and will result in low estimates of the potential for direct selection to shape female preference. Their study is therefore unable to resolve the question of how female preference for large forehead patches has evolved³.

Female preference for males with exaggerated ornaments is an important source of sexual selection⁴. For example, male collared flycatchers with enlarged forehead patches pair up with a female more rapidly than do normal males⁵. To resolve the controversy over whether such preferences can evolve as a by-product of selection acting on male ornaments³, Qvarnström *et al.*¹ investigate the quantitative genetics underlying variation in female mate choice in relation to male forehead patch size. Whereas patch size is moderately heritable, the heritability of female choice is close to zero, with a non-significant genetic correlation between the two. They argue that their estimated genetic correlation from a free-living bird population provides a better test of the potential for indirect selection than the strong genetic correlations found from laboratory studies on fish⁶ and insects⁷. We challenge this view, and draw attention to important limitations in how Qvarnström *et al.* measured mate choice.

Whereas laboratory studies use the power of

experimental choice trials, Qvarnström *et al.* use the patch size of a female's social partner. Believing that a female's social partner reflects her preference assumes that the distribution of male ornament size matches the distribution of female preference, and that every female can get what she wants. (This is akin to suggesting that Angelina Jolie is the only woman who finds Brad Pitt attractive and that most women prefer overweight men.) But in species such as the collared flycatcher, males and females are effectively removed from the pool of available partners when they form socially exclusive pair bonds. However, females mated to small-patched males do regularly copulate with extra-pair males with a larger patch, accounting for 90% of the sexual selection on male patch size⁸. By ignoring this important expression of preference, Qvarnström *et al.* overlook the main evolutionary pathway by which indirect selection on preference for large patches might act.

Given that mean female choice does not differ from mean male patch size and that variation in fitness is not larger in males than in females (Table 1 in ref. 1), the measure of choice used by Qvarnström *et al.* is unable to capture variation in female preference for large forehead patches. It therefore remains unclear whether variation in female preference affects which males females pair up with.

We agree with Qvarnström *et al.*¹ that partnership formation is shaped by a range of ecological processes, if only because not every female can partner with her preferred male. However, this reduces the importance of female preference in partnership formation *per se*, and determines both her indirect and direct fitness. The potential for both indirect and direct selection to shape the evolution of

preference is therefore, absolutely speaking, low. The issue about their relative importance cannot be settled until the potential for direct selection is quantified, and both are compared directly⁹. A non-significant genetic correlation between ornament and choice (between -0.3 and 0.3), and a significant heritability of choice (of 0.01 – 0.04) is not sufficient to support the argument.

Instead of dismissing results from the laboratory as unnatural, or from long-term wild data sets as unreplicated and correlational, we should study classic laboratory species in the wild and wild animals in captivity. This will allow for careful estimation of genetic variances and covariances of the appropriate traits, and thereby, for meaningful tests of evolutionary models of sexual selection¹⁰.

Erik Postma, Simon C. Griffith, Robert Brooks

School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales 2052, Australia
e-mail: e.postma@unsw.edu.au

1. Qvarnström, A., Brommer, J. E. & Gustafsson, L. *Nature* **441**, 84–86 (2006).
2. Kirkpatrick, M. & Ryan, M. J. *Nature* **350**, 33–38 (1991).
3. Kokko, H., Brooks, R., Jennions, M. D. & Morley, J. *Proc. R. Soc. Lond. B* **270**, 653–664 (2003).
4. Andersson, M. *Sexual Selection* (Princeton University Press, Princeton, New Jersey, 1994).
5. Qvarnström, A., Pärt, T. & Sheldon, B. C. *Nature* **405**, 344–347 (2000).
6. Bakker, T. C. M. *Nature* **363**, 255–257 (1993).
7. Iyengar, V. K., Reeve, H. K. & Eisner, T. *Nature* **419**, 830–832 (2002).
8. Sheldon, B. C. & Ellegren, H. *Anim. Behav.* **57**, 285–298 (1999).
9. Kirkpatrick, M. & Barton, N. H. *Proc. Natl Acad. Sci. USA* **94**, 1282–1286 (1997).
10. Charmantier, A. & Sheldon, B. C. *Trends Ecol. Evol.* **21**, 417–419 (2006).

doi:10.1038/nature05501

EVOLUTIONARY GENETICS

Qvarnström *et al.* reply

Replying to: E. Postma, S. C. Griffith & R. Brooks *Nature* **444**, doi: 10.1038/nature05501 (2006)

We have shown¹ that there is little scope for selection on male flycatchers' forehead patch size to drive the evolution of female choice for this ornament indirectly. Postma *et al.*² question this conclusion, arguing that a female's social partner (that is, realized mate choice) is not a good estimate of her preference, and that our estimates are biased because we do not take patterns of extra-pair paternity into account. However, indirect sexual selection can only operate through realized mate choice, and extra-pair copulations are associated with larger costs than indirect benefits.

Postma *et al.*² confuse the phenotypic and the genetic processes underlying indirect selection on mate choice. First, our use of mate choice is justified by the focus of models of indirect sexual selection on the heritability of realized preferences (that is, mate choice) and their assumption that a genetic correlation between ornament and choice is built up and maintained through a phenotypic correlation between mated pairs. A correlation between preference (unconstrained propensity to mate with a partner with a particularly sized ornament) and mate choice (its realized equivalent)

has been explicitly specified³, because preferences that are never realized cannot be exposed to selection, nor can their genes become associated with male genes for ornamentation. The point of Postma *et al.* that every female does not get what she wants in wild populations therefore underlines our conclusion of low potential for indirect sexual selection, rather than challenges it.

Second, the question is not whether one actress is the only person to find an actor attractive. It is whether her female relatives (who share her genes) make a mate choice that

is more similar to her's than to those made by other females in the population. Similarly, that the distribution of our measure of female mate choice of the ornament is equal to the distribution of the ornament does not falsify our definition of mate choice: the question is how the mate choices of different families are distributed (that is, how much of the variance in choice is due to genetic differences).

Extra-pair mating can be viewed as a less constrained form of mate choice than the choice of a social partner, but it is unlikely that taking into account extra-pair young, which constitute 15% of all offspring, would affect our conclusion that there is a low potential for indirect sexual selection. Engaging in extra-pair copulations is probably adaptive for males, although extra-pair copulations do not account for 90% of sexual selection on patch

size, because the study cited by Postma *et al.* (their ref. 8) does not compare all pathways of sexual selection (for example, pairing success is excluded), nor does it claim to. However, the crucial question is whether engaging in extra-pair copulations is adaptive for females: in flycatchers, and in many other passerines, the costs of extra-pair matings exceed the potential indirect benefits⁴, indicating that extra-pair copulation is not a major pathway through which indirect selection operates on female mate choice.

Laboratory studies explore the feasibility of processes involved in indirect sexual selection, but they cannot capture all the processes that together determine the strength of indirect sexual selection in natural populations. Studies like ours¹, which use long-term data on individually marked organisms in combination

with modern quantitative genetic techniques, offer such an opportunity.

Anna Qvarnström*, **Jon E. Brommer†**,
Lars Gustafsson*

*Department of Animal Ecology, Evolutionary Biology Centre, Uppsala University, 752 36 Uppsala, Sweden

e-mail: anna.qvarnstrom@ebc.uu.se

†Bird Ecology Unit, Department of Biological and Environmental Sciences, PO Box 65 (Viksbågen 1), University of Helsinki, 00014 Finland

1. Qvarnström, A., Brommer, J. E. & Gustafsson, L. *Nature* **441**, 84–86 (2006).
2. Postma, E., Griffith, S. C. & Brooks, R. *Nature* **444**, doi: 10.1038/nature05501 (2006).
3. Kirkpatrick, M. & Barton, N. H. *Proc. Natl Acad. Sci. USA* **94**, 1282–1286 (1997).
4. Arnqvist, G. & Kirkpatrick, M. *Am. Nat.* **165** (suppl.), 26–37 (2005).

doi:10.1038/nature05502

GEOCHEMISTRY

Biosignatures and abiotic constraints on early life

Arising from: Y. Ueno, K. Yamada, N. Yoshida, S. Maruyama & Y. Isozaki *Nature* **440**, 516–519 (2006)

Ueno *et al.*¹ contend that methane found in fluid inclusions within hydrothermally precipitated quartz in the Dresser Formation of western Australia (which is roughly 3.5 Gyr old) provides evidence for microbial methanogenesis in the early Archaean era. The authors discount alternative origins for this methane, suggesting that the range of $\delta^{13}\text{C}_{\text{CH}_4}$ values that they record (–56 to –36‰) is attributable to mixing between a primary microbial end-member with a $\delta^{13}\text{C}_{\text{CH}_4}$ value of less than –56‰ and a mature thermogenic gas enriched in ^{13}C (about –36‰). However, abiotic methane produced experimentally^{2,3} and in other Precambrian greenstone settings^{4,5} has ^{13}C -depleted $\delta^{13}\text{C}_{\text{CH}_4}$ values, as well as $\Delta^{13}\text{C}_{\text{CO}_2-\text{CH}_4}$ relationships that encompass the range measured for the inclusions by Ueno *et al.* — which suggests that an alternative, abiotic origin for the methane is equally plausible. The conclusions of Ueno *et al.* about the timing of the onset of microbial methanogenesis might not therefore be justified.

The range of isotopic values for abiotic methane shown in Fig. 2c of Ueno *et al.*¹ differs distinctly from those from thermogenic and microbial sources, and does not encompass the methane in their fluid inclusions. However, the range of $\delta^{13}\text{C}_{\text{CH}_4}$ for abiotic hydrocarbons is larger than that used in their comparison. Abiotic methane end-members in Precambrian Shield rocks extend to much lighter $\delta^{13}\text{C}_{\text{CH}_4}$ values (–47 to –28‰)^{4,5}. Also, as noted by Ueno *et al.*, experimental studies² indicate that abiotic synthesis using a Ni–Fe alloy catalyst may produce methane with a range of isotopic compositions ($\delta^{13}\text{C}_{\text{CH}_4}$ values of –54 to –19‰ at 200–300 °C) that is similar to the range in the inclusions¹. Although those

results have not been duplicated², they have not been disputed either. In fact, hydrothermal abiotic methane synthesis has been confirmed with Ni–Fe alloy⁶ and other potential mineral catalysts^{7,8}.

The values of $\Delta^{13}\text{C}_{\text{CO}_2-\text{CH}_4}$ measured by Ueno *et al.* in their inclusions (34–52‰; see their Supplementary Table 2) span a range that is identical to, or even smaller than, those produced in abiotic synthesis experiments (33–66‰)^{2,3}. Furthermore, the experimental range would encompass the larger value the authors infer for primary inclusions ($\Delta^{13}\text{C}_{\text{CO}_2-\text{CH}_4} > 52\%$). The authors also contend that the lack of C_2^+ compounds in those inclusions is in agreement with a microbial origin; however, C_2^+ hydrocarbons were not detected during methane synthesis in abiotic experiments² either, so such reactions can also produce methane with a low $\text{C}_2^+ / (\text{C}_1 + \text{C}_2^+)$ ratio (< 0.01). Therefore, neither an absence of C_2^+ nor isotopic compositions can distinguish between microbial and abiotic origins for this data set, even when both factors are considered together.

Ueno *et al.* also discount an abiotic source on the grounds that Fe–Ni alloys and other catalysts for Fischer–Tropsch-type synthesis reactions would not have existed in the surrounding rocks at conditions prevailing during silica precipitation. But abiotic methane could have been synthesized during interaction of migrating hydrothermal fluids with ultramafic rocks elsewhere in the system in the presence of either Ni–Fe alloy or other mineral catalysts, and transported to the site of quartz deposition^{3,9,10}, so the absence of suitable catalysts immediately adjacent to the fluid inclusion-bearing quartz would not preclude an abiotic origin. Indeed, Ueno *et al.* explain the

inferred thermogenic origin of methane in the secondary inclusions using migration from surrounding rocks.

The key criteria used by Ueno *et al.* to establish a microbial origin for methane in the inclusions from the Dresser Formation are therefore equally consistent with an abiotic origin, and their conclusion about the antiquity of the onset of microbial methanogenesis should be viewed with caution. Biosignatures that are mimicked by abiotic processes need to be analysed closely to evaluate not only signs of life in rocks from the early Earth but also the search for life on other planets.

Barbara Sherwood Lollar*,

Thomas M. McCollom†

*Department of Geology, University of Toronto, Toronto, Ontario M5S 3B1, Canada
e-mail: bslollar@chem.utoronto.ca

†Center for Astrobiology and Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado 80309, USA

1. Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S. & Isozaki, Y. *Nature* **440**, 516–519 (2006).
2. Horita, J. & Berndt, M. E. *Science* **285**, 1055–1057 (1999).
3. McCollom, T. M. & Seewald, J. S. *Earth Planet. Sci. Lett.* **243**, 74–84 (2006).
4. Sherwood Lollar, B., Westgate, T. D., Ward, J. A., Slater, G. F. & Lacrampe-Couloume, G. *Nature* **416**, 522–524 (2002).
5. Sherwood Lollar, B. *et al.* *Chem. Geol.* **226**, 328–339 (2006).
6. McCollom, T. M. & Seewald, J. S. *Geochim. Cosmochim. Acta* **67**, 3625–3644 (2003).
7. McCollom, T. M. & Seewald, J. S. *Geochim. Cosmochim. Acta* **65**, 3769–3778 (2001).
8. Foustoukos, D. I. & Seyfried, W. E. Jr *Science* **304**, 1002–1005 (2004).
9. Brasier, M. D. *et al.* *Precamb. Res.* **140**, 55–102 (2005).
10. Lindsay, J. F. *et al.* *Precamb. Res.* **143**, 1–22 (2005).

doi:10.1038/nature05499

GEOCHEMISTRY

Ueno *et al.* reply

Replying to: B. Sherwood Lollar & T. M. McCollom *Nature* **444**, doi: 10.1038/nature05499 (2006)

Sherwood Lollar and McCollom¹ suggest that abiotic reactions, in particular those catalysed by Fe–Ni alloys, might be responsible for the methane recorded in our fluid inclusions². They point out that ^{13}C in methane that is produced abiotically can be substantially depleted under field and laboratory conditions. However, native metal catalysis would not only have been implausible in the sulphidic environment of the basalt-hosted Dresser

hydrothermal system, but it would also have been inconsistent with the isotopic fractionation between CH_4 and CO_2 that we note in the primary fluid².

Sherwood Lollar and McCollom¹ overlook the petrography of the fluid inclusions, which shows that the $\delta^{13}\text{C}$ value of the primary methane should be less than –56‰ (ref. 2). However, this value is a maximum estimate of the fluid inclusion's $\delta^{13}\text{C}_{\text{CH}_4}$ value, because primary

methane would have been mixed with ^{13}C -enriched secondary methane². The isotopic fractionation between primary CH_4 and CO_2 should therefore be greater than 52‰.

Fischer–Tropsch-type reactions catalysed by Fe–Ni alloy at 200–300 °C (refs 3, 4) might produce methane depleted of ^{13}C by 30–50‰ relative to CO_2 ; even more isotopic fractionation between CH_4 and CO_2 is achieved at lower temperatures³. However, Fe

and Ni would have existed as sulphides at temperatures prevailing during silica precipitation ($<200^{\circ}\text{C}$)^{5–7}. In more than 300 rock samples from the Dresser Formation, we found that the primary mineral phases for Fe and Ni are mainly pyrite (FeS_2) and polydimite (Ni_3S_4), respectively⁶; these high-sulphur mineral assemblages cannot coexist with Fe–Ni alloy⁸ — in fact, we found no native metal in more than 300 petrographic thin sections.

The mineral assemblage suggests that conditions were highly sulphidic, which poisons Fischer–Tropsch-type reactions⁹, and that the primary inclusions contained hydrogen sulphide as well as methane². Metal oxides can catalyse these Fischer–Tropsch-type reactions¹, although they were present only as secondary minerals in the deposit we studied⁶. Hence, catalysis by native metals and their oxides would not be expected to produce primary methane in a low-temperature sulphidic environment.

On the basis of field studies^{10,11}, Sherwood Lollar and McCollom¹ suggest that Fe–Ni alloys and other catalysts might be present elsewhere in ultramafic rocks. But they overlook two important points. First, the Archaean hydrothermal fields they consider are the Apex Basalt⁸ and the Strelley Pool Chert⁹, which are above the Dresser Formation stratigraphically, and so rocks from the two different ages cannot give reliable information on the Dresser hydrothermal system. Second, no ultramafic rocks have been identified in the study area, despite thorough field mappings^{5–7,12,13}. The hydrothermal dykes that we studied occur in the uppermost 1,000 m of the pillowed basalt unit, which is more than 2,000 m thick^{5–7},

although the part deeper than the basalt unit is not exposed.

Native metal could have existed stably at depths greater than the site of silica precipitation, so methane could have been produced abiotically at much higher temperatures (exceeding 200°C). If so, isotopic fractionation between CO_2 and CH_4 should be smaller ($<30\%$ above 300°C , for example³). Deep and high-temperature abiotic synthesis does not therefore explain the observed isotopic fractionation of the primary fluid ($\delta^{13}\text{C}_{\text{CO}_2} - \delta^{13}\text{C}_{\text{CH}_4} \geq 52\%$).

Sherwood Lollar and McCollom¹ note that Precambrian Shield gas contains ^{13}C -depleted abiotic methane ($\delta^{13}\text{C}$ values between -47 and -28%)¹⁴, but this range is higher than that of the primary methane in our fluid inclusion (less than -56%). According to Sherwood Lollar and others¹⁴, the variation of $\delta^{13}\text{C}_{\text{CH}_4}$ values in Precambrian Shield gas is a result of mixing microbial methane (-60 to -55%) and abiotic methane (-45 to -25%). Therefore, a microbial origin for our ^{13}C -depleted primary methane (less than -56%) is not incompatible with their data or interpretation.

Consequently, metal-catalysed abiotic reactions would not explain the observed isotopic relationship. A microbial origin of the methane in hydrothermal dykes from the Dresser Formation (which is roughly 3.5 Gyr old) remains plausible.

Yuichiro Ueno*‡||, **Keita Yamada**§||, **Naohiro Yoshida***‡§||, **Shigenori Maruyama***†, **Yukio Isozaki**¶

*Research Center for the Evolving Earth and Planet, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8551, Japan
e-mail: yueno@depe.titech.ac.jp

†Department of Earth and Planetary Science, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8551, Japan

‡Department of Environmental Science and Technology, Tokyo Institute of Technology, Midori-ku, Yokohama 226-8503, Japan

§Department of Environmental Chemistry and Engineering, Tokyo Institute of Technology, Midori-ku, Yokohama 226-8503, Japan

||SORST project, Japan Science and Technology Corporation, Kawaguchi, Saitama 332-0012, Japan

¶Department of Earth Science and Astronomy, University of Tokyo, Meguro-ku, Tokyo 153-8902, Japan

1. Sherwood Lollar, B. & McCollom, T. M. *Nature* **444**, doi: 10.1038/nature05499 (2006).
2. Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S. & Isozaki, Y. *Nature* **440**, 516–519 (2006).
3. Horita, J. & Berndt, M. E. *Science* **285**, 1055–1057 (1999).
4. McCollom, T. M. & Seewald, J. S. *Earth Planet. Sci. Lett.* **243**, 74–84 (2006).
5. Ueno, Y., Isozaki, Y., Yurimoto, H. & Maruyama, S. *Int. Geol. Rev.* **43**, 196–212 (2001).
6. Ueno, Y., Yoshioka, H., Maruyama, S. & Isozaki, Y. *Geochim. Cosmochim. Acta* **68**, 573–589 (2004).
7. Kitajima, K., Maruyama, S., Utsunomiya, S. & Liou, J. G. *J. Metamorph. Geol.* **19**, 583–600 (2001).
8. Alt, J. C. & Shanks III, W. C. *J. Geophys. Res.* **103**, 9917–9930 (1998).
9. Anderson, R. B. in *Catalysis. Hydrocarbon Synthesis Hydrogenation and Cyclization* (ed. Emmett, P. H.) 29–255 (Reinhold, New York, 1956).
10. Brasier, M. D. et al. *Nature* **416**, 76–81 (2002).
11. Lindsay, J. F. et al. *Precamb. Res.* **143**, 1–22 (2005).
12. Nijman, W., de Bruijne, K. H. & Valkering, M. E. *Precamb. Res.* **95**, 247–274 (1999).
13. Van Kranendonk, M. J. *Earth Sci. Rev.* **74**, 197–240 (2006).
14. Sherwood Lollar, B. et al. *Chem. Geol.* **226**, 328–339 (2006).

doi:10.1038/nature05500

**Cover illustration**

Coloured scanning electron micrograph of adipocytes (pink) from bone-marrow tissue. (Courtesy of Science Photo Library.)

Editor, Nature

Philip Campbell

Insights Publisher

Sarah Greaves

Insights Editor

Lesley Anson

Commissioning Editor

Rory Howlett

Production Editor

Sarah Archibald

Senior Art Editor

Martin Harrison

Art Editor

Nik Spencer

Sponsorship

Claire Hines

Claudia Banks

Production

Susan Gray

Marketing

Katy Dunningham

Editorial Assistant

Laura Shaw

OBESITY AND DIABETES

Obesity and type 2 diabetes represent a serious threat to the health of the population of almost every country in the world. The World Health Organization estimates that 1.1 million people died as a result of diabetes in 2005, and this is almost certainly an underestimate. Moreover, the figure is expected to increase by 50% during the next ten years.

This escalation is due in part to increasing rates of childhood obesity. Until recently, type 2 diabetes was a disease that afflicted only adults. Now a growing number of children are being diagnosed with obesity-related type 2 diabetes. Also increasing is the incidence of 'metabolic syndrome' — a complex condition linked to obesity that is characterized by a cluster of closely related clinical features, including insulin resistance, dyslipidaemia and hypertension. Metabolic syndrome is associated with an increased risk of cardiovascular disease, which is ultimately responsible for a considerable proportion of diabetic mortality.

There is, therefore, an urgent need for new approaches to address obesity and type 2 diabetes and their associated complications. In particular, understanding the various processes that give rise to the characteristics of metabolic syndrome and its attendant risks — from abnormal regulation of energy metabolism through to dysfunction of molecular mechanisms — will pave the way for the development of new treatment strategies. And, as the reviews in this Insight highlight, it is hoped that a combination of preventative and therapeutic strategies will ultimately reduce the burden of type 2 diabetes on society.

We are pleased to acknowledge the support of Nestlé Research Center in producing this Insight. As always, *Nature* carries sole responsibility for all editorial content and peer review.

Deepa Nath, Marie-Thérèse Heemels and Lesley Anson, Senior Editors

REVIEWS

840 Mechanisms linking obesity to insulin resistance and type 2 diabetes

S. E. Kahn, R. L. Hull & K. M. Utzschneider

847 Adipocytes as regulators of energy balance and glucose homeostasis

E. D. Rosen & B. M. Spiegelman

854 Gut hormones and the regulation of energy homeostasis

K. G. Murphy & S. R. Bloom

860 Inflammation and metabolic disorders

G. S. Hotamisligil

868 Sirtuins as potential targets for metabolic syndrome

L. Guarente

875 Mechanisms linking obesity with cardiovascular disease

L. F. Van Gaal, I. L. Mertens & C. E. De Block

881 Abdominal obesity and metabolic syndrome

J.-P. Després & I. Lemieux

nature
insight

Mechanisms linking obesity to insulin resistance and type 2 diabetes

Steven E. Kahn¹, Rebecca L. Hull¹ & Kristina M. Utzschneider¹

Obesity is associated with an increased risk of developing insulin resistance and type 2 diabetes. In obese individuals, adipose tissue releases increased amounts of non-esterified fatty acids, glycerol, hormones, pro-inflammatory cytokines and other factors that are involved in the development of insulin resistance. When insulin resistance is accompanied by dysfunction of pancreatic islet β -cells — the cells that release insulin — failure to control blood glucose levels results. Abnormalities in β -cell function are therefore critical in defining the risk and development of type 2 diabetes. This knowledge is fostering exploration of the molecular and genetic basis of the disease and new approaches to its treatment and prevention.

The increased prevalence of obesity has focused attention on a world-wide problem that is not one of famine or infection, but one of surplus. In the United States, only about a third of adults are considered to be of 'normal' weight¹, and similar trends are being observed worldwide². Obesity is associated with several conditions, the most devastating of which may be type 2 diabetes. At the turn of this century 171 million individuals were estimated to have diabetes, and this is expected to increase to 366 million by 2030 (ref. 3).

Both obesity and type 2 diabetes are associated with insulin resistance⁴. But most obese, insulin-resistant individuals do not develop hyperglycaemia. Under normal conditions, the pancreatic islet β -cells increase insulin release sufficiently to overcome the reduced efficiency of insulin action, thereby maintaining normal glucose tolerance^{5–7}. For obesity and insulin resistance to be associated with type 2 diabetes, β -cells must be unable to compensate fully for decreased insulin sensitivity⁸. β -cell dysfunction exists in individuals who are at high risk of developing the disease even when their glucose levels are still normal⁸. Non-esterified fatty acids (NEFAs) induce insulin resistance and impair β -cell function, making them a likely culprit.

Here we explore the interactions between obesity, insulin resistance and β -cell dysfunction that result in type 2 diabetes, and attempt to lay a framework to consider approaches that might lessen the burden of these epidemics that threaten modern society.

Insulin resistance and obesity

Fluctuations in insulin sensitivity occur during the normal life cycle, with insulin resistance being observed during puberty⁹ and pregnancy¹⁰, and with ageing¹¹. Conversely, lifestyle variation such as increased physical activity¹² and increased carbohydrate intake¹³ are associated with enhanced insulin sensitivity.

The most critical factor in the emergence of metabolic diseases is obesity. Adipose tissue modulates metabolism by releasing NEFAs and glycerol, hormones — including leptin and adiponectin — and proinflammatory cytokines^{14–16}. In obesity, the production of many of these products is increased. Retinol-binding protein-4 (RBP4) induces insulin resistance through reduced phosphatidylinositol-3-OH kinase (PI(3)K) signalling in muscle and enhanced expression of the gluconeogenic enzyme phosphoenolpyruvate carboxykinase in the liver through

a retinol-dependent mechanism¹⁷. By contrast, adiponectin acts as an insulin sensitizer, stimulating fatty acid oxidation in an AMP-activated protein kinase (AMPK) and peroxisome proliferator activated receptor- α (PPAR- α)-dependent manner^{15,18}.

In addition to adipocyte-derived factors, increased release of tumour necrosis factor- α (TNF- α), interleukin-6 (IL-6), monocyte chemo-attractant protein-1 (MCP-1) and additional products of macrophages and other cells that populate adipose tissue might also have a role in the development of insulin resistance^{14,19}. TNF- α and IL-6 act through classical receptor-mediated processes to stimulate both the c-Jun amino-terminal kinase (JNK) and the I κ B kinase- β (IKK- β)/nuclear factor- κ B (NF- κ B) pathways, resulting in upregulation of potential mediators of inflammation that can lead to insulin resistance.

Pathways involving the induction of suppression of cytokine signalling (SOCS) proteins²⁰ and inducible nitric oxide synthase (iNOS)²¹ may be involved in mediating cytokine-induced insulin resistance. Secretion of these proinflammatory proteins, particularly MCP-1 by adipocytes, endothelial cells and monocytes, increases macrophage recruitment and thereby contributes to a feedforward process^{22,23}.

The release of NEFAs may be the single most critical factor in modulating insulin sensitivity. Increased NEFA levels are observed in obesity and type 2 diabetes, and are associated with the insulin resistance observed in both^{24,25}. Insulin resistance develops within hours of an acute increase in plasma NEFA levels in humans²⁶. Conversely, insulin-mediated glucose uptake and glucose tolerance improve with an acute decrease in NEFA levels after treatment with the antilipolytic agent acipimox²⁷. Increased intracellular NEFAs might result in competition with glucose for substrate oxidation leading to the serial inhibition of pyruvate dehydrogenase, phosphofructokinase and hexokinase II activity²⁸. It has also been proposed that increased NEFA delivery or decreased intracellular metabolism of fatty acids results in an increase in the intracellular content of fatty acid metabolites such as diacylglycerol (DAG), fatty acyl-coenzyme A (fatty acyl-CoA), and ceramides, which, in turn, activate a serine/threonine kinase cascade leading to serine/threonine phosphorylation of insulin receptor substrate-1 (IRS-1) and insulin receptor substrate-2 (IRS-2), and a reduced ability of these molecules to activate PI(3)K²⁹. Subsequently, events downstream of insulin-receptor signalling are diminished.

¹Division of Metabolism, Endocrinology and Nutrition, Department of Medicine, VA Puget Sound Health Care System and University of Washington, 1660 South Columbian Way, Seattle, Washington 98108, USA.

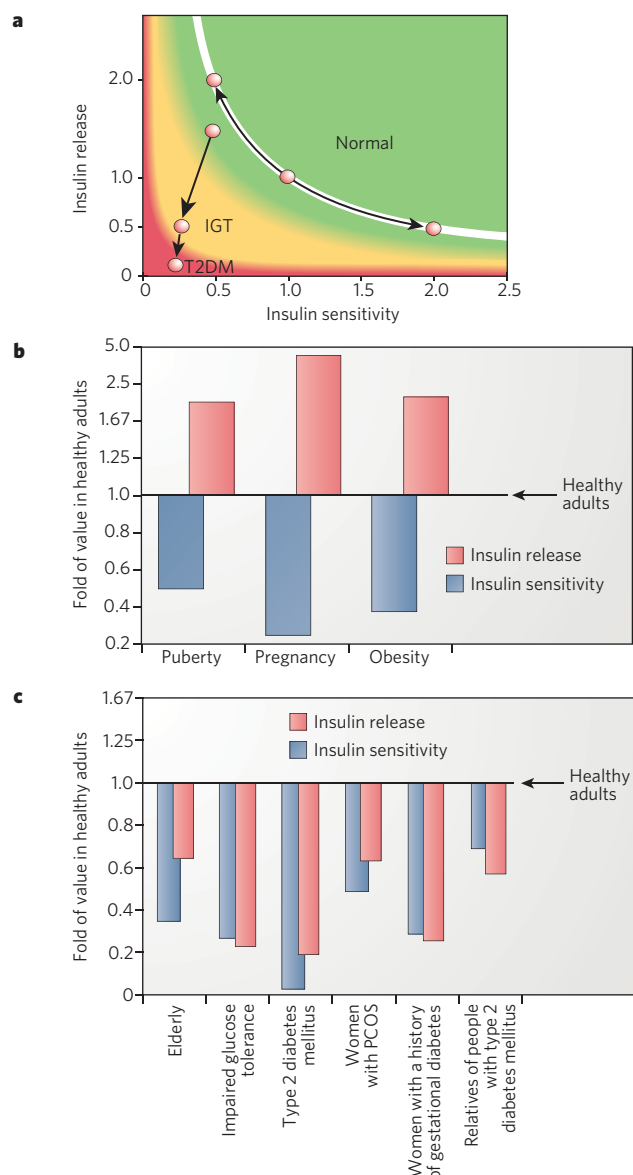


Figure 1 | Relationship between insulin sensitivity and insulin release in health and disease. **a**, Relationship between insulin sensitivity and the β -cell insulin response is nonlinear. This hyperbolic relationship means that assessment of β -cell function requires knowledge of both insulin sensitivity and the insulin response. Hypothetical regions delineating normal glucose tolerance (green), impaired glucose tolerance (IGT; yellow) and type 2 diabetes mellitus (T2DM; red) are shown. In response to changes in insulin sensitivity, insulin release increases or decreases reciprocally to maintain normal glucose tolerance — ‘moving up’ or ‘moving down’ the curve. In individuals who are at high risk of developing type 2 diabetes, the progression from normal glucose tolerance to type 2 diabetes transitions through impaired glucose tolerance and results in a ‘falling off the curve’. Those individuals who do progress will frequently have deviated away from the curve even when they have normal glucose tolerance, in keeping with β -cell function already being decreased before the development of hyperglycaemia. **b**, Insulin sensitivity and insulin responses during puberty, during pregnancy and in obesity relative to that in healthy adults. On the basis of the hyperbolic relationship defining β -cell function, the product of insulin sensitivity and the insulin response is 1. In individuals with normal β -cells, glucose tolerance is preserved during puberty, during pregnancy and in obesity as the decrease in insulin sensitivity is matched by a reciprocal, compensatory increase in insulin release, maintaining the product of 1. **c**, Insulin sensitivity and insulin responses in groups of people with type 2 diabetes and those at increased risk of developing type 2 diabetes. In these groups, the decline in insulin sensitivity is not matched by a reciprocal increase in the insulin response. Instead, the insulin response also declines so the product is less than 1, which is compatible with the idea of β -cell dysfunction. PCOS, polycystic ovarian syndrome.

The distribution of body fat is itself a critical determinant of insulin sensitivity. Whereas simple obesity is typically associated with insulin resistance, insulin sensitivity also varies markedly in lean individuals because of differences in body fat distribution^{30–33}. Lean individuals with a more peripheral distribution of fat are more insulin sensitive than lean subjects who have their fat distributed predominantly centrally — that is, in the abdominal and chest areas.

Differences in the characteristics of adipose tissue from these two depots might explain in part why the metabolic effects of intra-abdominal and subcutaneous fat differ. For example, intra-abdominal fat expresses more genes encoding secretory proteins and proteins responsible for energy production³⁴. The amount of protein released per adipocyte also differs according to their location^{19,35}. The secretion of adiponectin by omental adipocytes is greater than that of subcutaneous-derived adipocytes, and the amount released from these omental adipocytes is more strongly and negatively correlated with body mass index (BMI)³⁵. Small adipocytes release more adiponectin than do larger cells, and omental adipocytes are typically smaller than subcutaneous fat cells³⁶. Although each adipocyte from the intra-abdominal depot secretes more adiponectin, the subcutaneous depot represents a greater proportion of total body fat, and thus its contribution to total adiponectin levels will invariably be greater.

The delivery of NEFAs to the tissues might also be modulated by their source. Intra-abdominal fat is more lipolytic than subcutaneous fat and is also less sensitive to the anti-lipolytic effect of insulin³⁷. This difference in adipocyte characteristics, combined with the proximity of the liver to the intra-abdominal fat depot, probably results in greater exposure of this organ than the peripheral tissues to NEFAs. This difference in exposure and the presence of a portal–peripheral NEFA gradient could explain why the liver can be insulin resistant at a time when the peripheral tissues are not³⁸.

β -cell function and mass

β -cells are markedly plastic in their ability to regulate insulin release, but at the same time do so in a very precise manner. The quantity of insulin released by β -cells varies according to the nature, quantity and route of administration of the stimulus, and the prevailing glucose concentration. In this manner, the β -cell is crucial to ensuring that in healthy subjects plasma glucose concentrations remain within a relatively narrow physiological range.

Insulin sensitivity also modulates β -cell function and is almost always decreased in obesity. Insulin-resistant individuals, whether lean or obese, have greater insulin responses and lower hepatic insulin clearance than insulin-sensitive individuals. In healthy individuals, there is a feedback loop between the insulin-sensitive tissues and the β -cells, with β -cells increasing insulin supply in response to demand by the liver, muscles and adipose tissue⁷. The relationship between insulin sensitivity and insulin levels is reciprocal and nonlinear in nature (Fig. 1a). In order for glucose tolerance to remain unchanged, changes in insulin sensitivity must be matched by a proportionate yet opposite change in circulating insulin levels. Failure of this feedback loop results in a deviation from normal glucose tolerance and underlies the development of diabetes.

Another important implication of this feedback loop is that interpretation of the β -cell's secretory response to a given stimulus must take into account the prevailing degree of insulin sensitivity. This ability of the β -cell to adapt to changes in insulin sensitivity seems to result from two parameters: the functional responsiveness of the cell and β -cell mass. In response to the insulin resistance observed in obesity^{5–7}, puberty⁹ and pregnancy¹⁰, human β -cells can increase insulin release to levels fourfold to fivefold higher than in insulin-sensitive individuals (Fig. 1b), whereas β -cell volume is only enhanced by about 50%^{39,40}.

The integration of the β -cell's response to changes in insulin sensitivity probably involves increased cellular glucose metabolism, NEFA signalling and sensitivity to incretins (Fig. 2). Increased β -cell glucose metabolism occurs in animal models of obesity that maintain normal blood glucose levels (euglycaemia)⁴¹. Glucose-stimulated insulin secretion requires the metabolism of glucose and thereby the generation of

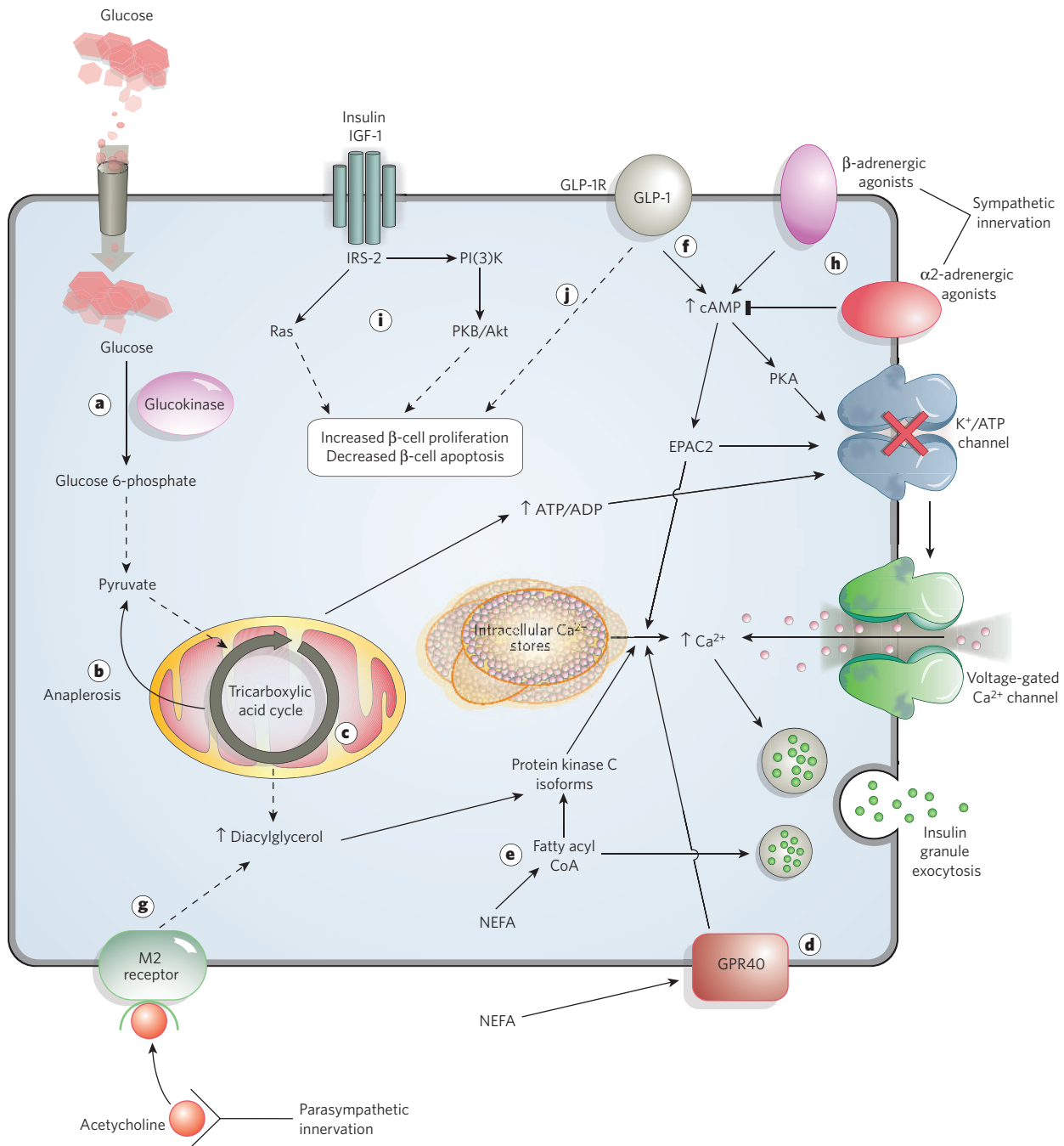


Figure 2 | Simplified model outlining potential cellular mechanisms of β -cell adaptation to insulin resistance. Glucose-stimulated insulin secretion occurs by oxidative metabolism of glucose, leading to an increase in the ATP/ADP ratio. This causes closure of K^+_{ATP} channels, depolarization of the plasma membrane, increased cytoplasmic calcium concentrations through voltage-gated calcium channels, and exocytosis of insulin-containing secretory granules. In conditions in which insulin demand is increased, β -cell glucose metabolism can be enhanced by increased glucokinase enzyme activity (a) and by replenishment of tricarboxylic acid cycle intermediates by anaplerosis (b). Glucose-induced increases in citrate levels lead to increased amounts of malonyl CoA (c), which, through inhibition of carnitine palmitoyl transferase-1 (CPT1), leads to increased levels of long-chain acyl CoA, increased diacylglycerol (DAG) and signalling through protein kinase C (PKC). Fatty acids influence insulin release by signalling through the G-protein-coupled receptor GPR40 (d) or through metabolism to fatty acyl CoA (e) and stimulation of insulin granule exocytosis, either directly or through PKC-dependent mechanisms.

The incretin GLP-1 potentiates glucose-stimulated insulin release through its G-protein-coupled receptor (f) by means of mechanisms that include stimulation of protein kinase A (PKA) and the guanine nucleotide exchange factor EPAC2. Release of acetylcholine from parasympathetic nerve terminals activates the M2 muscarinic receptor (g), stimulating insulin release in a DAG- and PKC-dependent manner. Dual actions on insulin secretion have been described for sympathetic nerves (h), with α 2-adrenergic agonists inhibiting and β -adrenergic agonists stimulating insulin secretion. Both pathways act through adenylyl cyclase, resulting in a decrease or increase in cAMP levels, respectively. β -cell mass can be positively regulated by the insulin/IGF-1 receptor signalling pathway (i) in which IRS-2 becomes phosphorylated, activating a cascade of downstream molecules including PI(3)K and PKB/Akt and Ras, resulting in enhanced β -cell survival. Finally, GLP-1 receptor (GLP-1R) signalling can similarly enhance β -cell survival and inhibit β -cell apoptosis (j) through several pathways, including transactivation of the epidermal growth factor receptor and stimulation of the IRS-2 pathway.

ATP. The resulting increase in the ATP/ADP ratio triggers the closure of the ATP-sensitive potassium (K^+_{ATP}) channel, depolarization of the cell membrane and influx of calcium through voltage-dependent calcium channels, resulting in insulin granule exocytosis. The increase in β -cell glucose metabolism involves an increase in the activity of glucokinase, the rate-limiting enzyme responsible for glucose phosphorylation after its entry into the cell⁴¹ (Fig. 2a). There is evidence that glucose use rises as both oxidation and flux of glucose are increased, the latter through pyruvate carboxylase and the replenishment of tricarboxylic acid cycle intermediates in the mitochondria, a process known as anaplerosis⁴² (Fig. 2b).

Increased citrate levels generated by glucose metabolism lead to generation of malonyl-CoA and increased long-chain acyl-CoA and diacylglycerol levels through inhibition of carnitine palmitoyl transferase 1. This leads to protein kinase C (PKC) activation and stimulation of insulin release (Fig. 2c). Despite these animal data, studies in humans suggest that increased glucose levels are not responsible for the adaptive increase in insulin release in response to decreased insulin sensitivity. For example, experimental insulin resistance was associated with increased insulin release both in the fasting state and after stimulation, yet the fasting plasma glucose level did not increase⁴³. Improving insulin sensitivity by exercise training resulted in the expected fall in insulin levels, but the fasting glucose level was in fact higher after the intervention⁴⁴.

NEFAs are important for normal β -cell function, and potentiate insulin release in response to glucose and non-glucose secretagogues^{45,46}. Studies in dogs suggest that it is the nocturnal elevations in NEFAs that might underlie the β -cell's adaptive response to insulin resistance⁴⁷. This might involve two different mechanisms. The first relates to the binding of NEFAs to the G-protein-coupled receptor GPR40 on the cell membrane, resulting in the activation of intracellular signalling and a subsequent increase in intracellular calcium and secretory granule exocytosis⁴⁸ (Fig. 2d). The other involves generation of fatty acyl-CoA (Fig. 2e), which increases insulin release both by directly stimulating secretory granule exocytosis and by PKC activation⁴⁶.

A third possible mechanism, whereby a humoral factor could mediate increased β -cell output in response to decreased insulin sensitivity, is increased sensitivity to incretin hormones. These are produced in the intestinal mucosa and are responsible for the enhancement of the insulin response observed after oral — compared with intravenous — glucose administration⁴⁹. Plasma levels of glucagon-like peptide-1 (GLP-1; Fig. 2f) are not increased in obese individuals and are, in fact, decreased in individuals who are morbidly obese⁵⁰. However, the insulin response to nutrient ingestion in these individuals is increased, suggesting that, as with glucose and NEFAs, the β -cell might become more responsive to the effects of this peptide to modulate insulin secretion.

The extensive innervation of the islet by both parasympathetic and sympathetic neurons, and the intimate involvement of the central nervous system (CNS) in the regulation of metabolism suggest that the CNS might also have an important role in the functional adaptation to changes in insulin sensitivity. Increased insulin release is observed immediately after experimental lesioning of the ventromedial hypothalamus (VMH) and this effect is mediated by increased vagal activity, which can be blocked by vagotomy⁵¹. Parasympathetic stimulation of insulin release occurs through activation by acetylcholine of the M2 muscarinic receptor on the β -cell surface (Fig. 2g). The sympathetic nervous system is also important, with increased activity of the α 2-adrenergic component being associated with decreased insulin release, whereas increased β -adrenergic activity enhances insulin output⁵² (Fig. 2h).

Although changes in β -cell function are observed under conditions of increased secretory demand, the volume of β -cells also increases. In rodents fed a high-fat diet for 12 months to induce obesity and insulin resistance, islet size increases as a result of an increase in the number of β -cells rather than a change in β -cell size, and new islets do not form⁵³. Pregnancy, and its accompanying insulin resistance, is associated with β -cell proliferation in rodents⁵⁴. Human studies suggest that β -cell volume is increased by about 50% in healthy obese individuals, but that this

increase might be more dependent on hypertrophy of existing cells than proliferation. Any formation of new β -cells probably occurs as a result of new islet formation derived from pancreatic exocrine ducts^{39,40}.

Glucose and/or NEFAs might mediate the increase in β -cell mass. In rats, infusion of glucose for up to 4 days results in β -cell hyperplasia and hypertrophy, together with a concomitant increase in glucose-stimulated insulin secretion^{55,56}. But we failed to observe an increase in fed or fasted glucose levels during 12 months of increased dietary fat feeding, suggesting that glucose is unlikely to be the mediator of the observed increase in β -cell mass in this model⁵³. By contrast, infusion of lipid for 4 days results in increased β -cell mass in rats owing to β -cell proliferation, with no sustained improvement in insulin release⁵⁶. This is more consistent with our 12-month increased dietary fat feeding study, in which β -cell mass increased but glucose-induced insulin release did not, representing a disassociation between β -cell mass and secretory function.

Increased signalling by insulin and/or insulin-like growth factor 1 (IGF-1) is a potentially important pathway for modulation of islet mass. Activation of the insulin/IGF-1 receptor leads to phosphorylation of IRS-2 and downstream signalling through pathways including PI(3)K/protein kinase-B (PKB/Akt) and Ras, leading to activation of the mitogen-activated protein (MAP) kinases ERK-1 and ERK-2 (ref. 57; Fig. 2i). Increased β -cell expression of IRS-2 is associated with increased β -cell proliferation, neogenesis and survival⁵⁸. In *Irs-2*-knockout mice, reintroduction of β -cell IRS-2 alone reversed the diabetic phenotype by increasing β -cell replication, resulting in an increase in β -cell number and mass⁵⁹.

The incretin GLP-1 is an insulin secretagogue but is also a β -cell mitogen, capable of increasing β -cell proliferation and reducing β -cell apoptosis in animal models⁴⁹ (Fig. 2j). Whether GLP-1 has similar effects in humans is not known.

Neural signalling might also regulate β -cell mass. Experimental lesions of the VMH produce a model of obesity and insulin resistance that is associated with vagal hyperactivity and proliferation of islet cells, particularly β -cells⁶⁰. Thus, it is possible that increased vagal input associated with diet-induced obesity might also contribute to increased β -cell mass.

β -cell dysfunction

When the β -cell is healthy, the adaptive response to insulin resistance involves changes in both function and mass, and is so efficient that normal glucose tolerance is maintained. But when β -cell dysfunction is present, impaired glucose tolerance, impaired fasting glucose and, at the extreme, type 2 diabetes result.

The magnitude of the reduction in β -cell function in type 2 diabetes is compatible with a failure of the cell to respond adequately to secretagogue stimulation, an important contributor to reduced insulin release. This conclusion is based on a number of observations. First, the β -cell is unable to release insulin rapidly in response to intravenous glucose, despite the fact that the β -cells in individuals with type 2 diabetes clearly contain insulin⁸. Second, delivery of non-glucose secretagogues can acutely increase insulin release but does not result in equivalent responses to those seen with similar stimulation in healthy subjects⁸. Third, although the number of β -cells is clearly reduced by about 50% in type 2 diabetes^{39,40}, this degree of β -cell loss cannot fully account for the change in secretory function, because by the time the diagnostic level for diabetes occurs, the cell is operating at 25% or less of its functional capacity⁶¹.

Type 2 diabetes is progressive, and one of the main factors responsible for this is a continued decline in β -cell function⁸. As a result of β -cell dysfunction and inadequate insulin secretion, postprandial and subsequently fasting glucose levels increase owing to incomplete suppression of hepatic glucose production and decreased efficiency of liver and muscle glucose uptake. The extremely elevated blood glucose levels frequently observed in diabetes might contribute to further disease progression through glucotoxic effects on the β -cell and harmful effects on insulin sensitivity, both of which can be ameliorated by therapeutically lowering the glucose level⁶². By contrast, raising the blood glucose

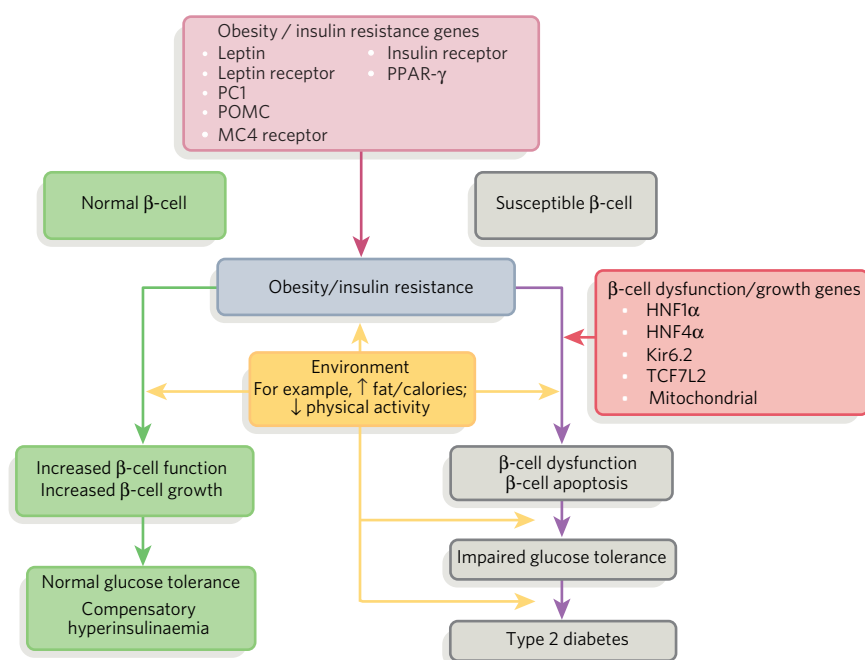


Figure 3 | Interaction of genes and the environment in individuals who maintain normal glucose tolerance and those who develop type 2 diabetes. Genes responsible for obesity and insulin resistance interact with environmental factors (increased fat/caloric intake and decreased physical activity), resulting in the development of obesity and insulin resistance. These increase secretory demand on β -cells. If the β -cells are normal, their function and mass increase in response to this increased secretory demand, leading to compensatory hyperinsulinaemia and the maintenance of normal glucose tolerance. By contrast, susceptible β -cells have a genetically determined risk, and the combination of increased secretory demand and detrimental environment result in β -cell dysfunction and decreased β -cell mass, resulting in progression to impaired glucose tolerance, followed, ultimately, by the development of type 2 diabetes. HNF, hepatocyte nuclear factor.

level for 20 hours in healthy subjects has exactly the opposite effect: it improves insulin sensitivity and enhances β -cell function⁶³. This suggests that a pre-existing, and perhaps genetically determined, risk is crucial for β -cell dysfunction to occur. It is this pre-existing abnormality that results, with time, in a progressive impairment in insulin release and, ultimately, an increase in glucose levels, the latter of which further aggravates the situation and thereby contributes to β -cell failure.

A second metabolic derangement that might contribute in a feedforward manner to progressive loss of β -cell function is elevated plasma NEFA concentrations. Although NEFAs are critical for normal insulin release, chronic exposure to NEFAs *in vitro* and *in vivo* is associated with marked impairments in glucose-stimulated insulin secretion and decreased insulin biosynthesis^{64,65}. Elevated NEFA levels produced by a lipid infusion *in vivo* contribute to the development of insulin resistance and also prevent the expected compensatory β -cell response in humans⁶⁶. This dual effect makes them a good candidate to link insulin resistance and β -cell dysfunction in individuals with type 2 diabetes and those at risk of the disorder. This lipotoxic effect can also act synergistically with glucose to produce even greater deleterious effects, commonly referred to as 'glucolipotoxicity'.

Pathogenesis of type 2 diabetes

Given that β -cell function is decreased by about 75% when fasting hyperglycaemia is present, assessment of β -cell function in individuals at risk of developing diabetes has been of interest. Even when the glucose level is still within the normal range, β -cell function decreases progressively as the fasting glucose level increases⁶⁷. Groups at increased risk of subsequently developing diabetes exhibit β -cell dysfunction well before they would be considered to have reduced glucose tolerance, in keeping with the idea of a pre-existing risk. Examples include women with a history of gestational diabetes⁶⁸ or polycystic ovarian syndrome⁶⁹, older subjects, who frequently develop hyperglycaemia as they continue to age⁷⁰, and individuals with impaired glucose tolerance^{71,72} (Fig. 1c). First-degree relatives of individuals with type 2 diabetes, who are genetically at increased risk, also have impaired β -cell function, even though they may still have normal glucose tolerance⁷³ (Fig. 1c). Data from groups of first-degree relatives with different ethnic backgrounds highlight that common processes underlie the development of type 2 diabetes — namely insulin resistance and β -cell dysfunction — with the degree of abnormality of insulin release being the dominant determinant of differences in glucose tolerance between individuals⁷².

Longitudinal data examining the progression to type 2 diabetes have been collected from the Pima Indians, in whom the prevalence of diabetes is higher than almost any other group in the world. In these individuals, who are insulin resistant, the transition from normal to impaired glucose tolerance and then on to diabetes is characterized by a progressive loss of β -cell function⁷⁴. Those who did not progress to diabetes over time simply increased their insulin output as insulin sensitivity declined. In those individuals who progressed, the presence of a defect in insulin release was already manifest at their initial assessment, even though at that time there was nothing else to indicate that they would ultimately develop diabetes. These findings have been confirmed for non-Hispanic whites, African Americans and Hispanics participating in the Insulin Resistance Atherosclerosis Study (IRAS)⁷⁵.

Genes and environment

Many genes interact with the environment to produce obesity and diabetes (Fig. 3). In the case of obesity, the most frequent mutation is that in the melanocortin-4 receptor, which accounts for up to 4% of cases of severe obesity. Other rare causes include mutations in leptin and the leptin receptor, prohormone convertase 1 (PC1) and pro-opiomelanocortin (POMC)⁷⁶. The gene variant most commonly associated with insulin sensitivity is the P12A polymorphism in *PPAR* γ , which is associated with an increased risk of developing diabetes^{77,78}. A number of genes associated with β -cell dysfunction have been identified, and include hepatocyte nuclear factor-4 α and 1 α — genes known to cause the monogenic disorder maturity onset diabetes of the young (MODY) — the E23K polymorphism in the islet ATP-sensitive potassium channel Kir6.2 (encoded by *KCNJ11*), two non-coding single-nucleotide polymorphisms in transcription factor 7-like 2 (*TCF7L2*) and mutations in the mitochondrial genome that are also associated with neurosensory hearing loss⁷⁷. Work is ongoing on many candidate genes, including calpain 10, adiponectin, *PPAR* γ coactivator 1 (*PGC1*) and the glucose transporter *GLUT2* (ref. 77).

Environmental factors are largely responsible for the modern day epidemic of obesity and type 2 diabetes. Increased caloric availability and fat consumption in the setting of decreased physical activity lead to over-nutrition, increased nutrient storage and obesity. Long-term increased dietary fat intake is associated not only with the development of obesity but also with reductions in insulin release⁷⁹. This effect has important consequences if β -cell function is already inherently abnormal owing to genetic susceptibility. Furthermore, changes in the proportions of dietary carbohydrate and fat can affect both insulin sensitivity and insulin release

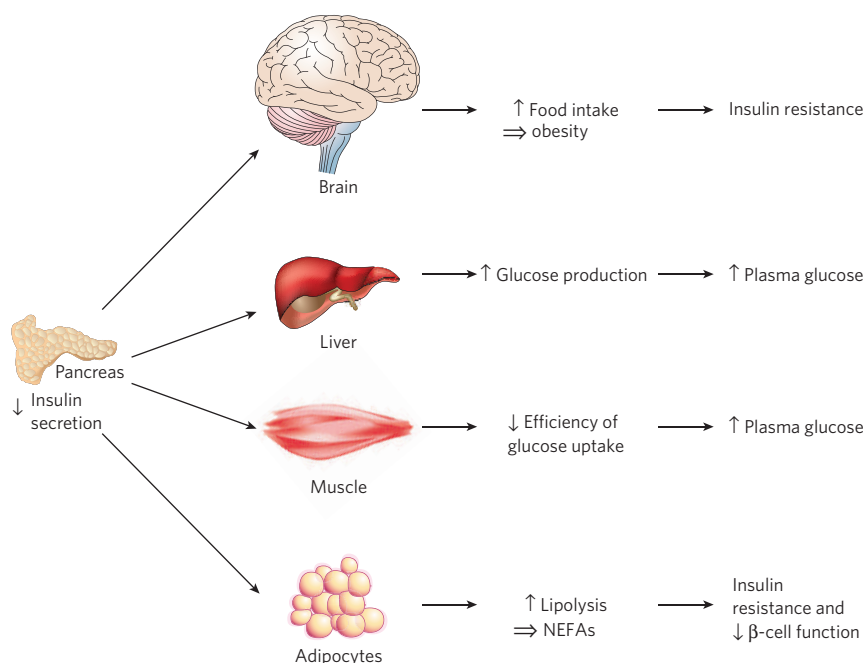


Figure 4 | Model of the critical role of impaired insulin release in linking obesity with insulin resistance and type 2 diabetes. Impaired insulin secretion results in decreased insulin levels and decreased signalling in the hypothalamus, leading to increased food intake and weight gain, decreased inhibition of hepatic glucose production, reduced efficiency of glucose uptake in muscle, and increased lipolysis in the adipocyte, resulting in increased plasma NEFA levels. The increase in body weight and NEFAs contribute to insulin resistance, and the increased NEFAs also suppress the β -cell's adaptive response to insulin resistance. The increased glucose levels together with the elevated NEFA levels can synergize to further adversely affect β -cell health and insulin action, often referred to as 'glucolipotoxicity'.

within three days of changing nutrient balance, at a time when obesity would not yet be a factor¹³. Another proposed environmental mechanism is thought to occur *in utero* and/or during the early postnatal period when poor nutrition alters metabolism, resulting in a tissue adaptation that favours the storage of nutrients⁸⁰. The end result of these environmental changes is a deleterious interaction with genes that predispose to the development of obesity and type 2 diabetes.

A possible unifying mechanism

Having a single mechanism to explain the link between obesity, insulin resistance and type 2 diabetes would be ideal. A defect in insulin release by the β -cell could be crucial (Fig. 4). Decreased insulin release could result in disordered regulation of glucose levels by decreasing suppression of hepatic glucose production and reducing the efficiency of glucose uptake in insulin-sensitive tissues. Decreased insulin output could also impair adipocyte metabolism, resulting in increased lipolysis and elevated NEFA levels. Elevations in both NEFAs and glucose can occur simultaneously, and together are more deleterious to islet health and insulin action than either alone^{46,81}. Thus the process may slowly feed forward, in keeping with observations that the onset of type 2 diabetes is usually a slow process that takes many years.

Even mild impairments of insulin release may have central effects on metabolic homeostasis. Insulin acts in the hypothalamus to regulate body weight, and impaired insulin signalling is associated with changes in food intake and body weight⁸². Thus, β -cell dysfunction resulting in a relative reduction in insulin release would be expected to result in decreased insulin action in this crucial brain region and be associated with weight gain and an aggravation of insulin resistance.

Insulin resistance at the level of the β -cell might have a role in the pathogenesis of defective insulin release. This idea is based mainly on studies using mice with a β -cell selective deletion of the insulin receptor and by subsequent work in which key molecules in the insulin signalling cascade within the β -cell have been manipulated^{57,83}. However, deletion of the insulin receptor in the β -cell might also result in a loss of insulin receptors in the hypothalamus, so it is not clear whether the resultant effects are centrally mediated or truly β -cell specific. Although there is currently no evidence that insulin receptor mutations are commonly associated with type 2 diabetes, a reduction in insulin signalling in the β -cell remains an interesting possibility in further integrating defects in insulin action into the pathogenesis of obesity and type 2 diabetes.

Future directions

The past decade has seen major advances in our understanding of the relationship between obesity, insulin resistance and type 2 diabetes. However, despite these tremendous strides and the identification of the critical nature of β -cell dysfunction in the development of type 2 diabetes, there is still a great deal to be learned about the mechanisms linking obesity, insulin resistance and type 2 diabetes. Although clinical studies aimed at reducing the deleterious effects of these conditions have been undertaken and more are being launched, a better understanding of the genetic bases of these processes and the cellular events that underlie them should enhance our ability to devise new and better approaches to try to stem the deleterious effects of these diseases.

- Hedley, A. A. *et al.* Prevalence of overweight and obesity among US children, adolescents, and adults, 1999–2002. *J. Am. Med. Assoc.* **291**, 2847–2850 (2004).
- World Health Organization Consultation on Obesity 1–253 (World Health Organization, Geneva, 2000).
- Wild, S., Roglic, G., Green, A., Sicree, R. & King, H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* **27**, 1047–1053 (2004).
- Reaven, G. M. Role of insulin resistance in human disease. *Diabetes* **37**, 1595–1607 (1988).
- Perley, M. & Kipnis, D. M. Plasma insulin responses to glucose and tolbutamide of normal weight and obese diabetic and nondiabetic subjects. *Diabetes* **15**, 867–874 (1966).
- Polonsky, K. S., Given, B. D. & Van Cauter, E. Twenty-four-hour profiles and patterns of insulin secretion in normal and obese subjects. *J. Clin. Invest.* **81**, 442–448 (1988).
- Kahn, S. E. *et al.* Quantification of the relationship between insulin sensitivity and B-cell function in human subjects. Evidence for a hyperbolic function. *Diabetes* **42**, 1663–1672 (1993).
- Kahn, S. E. The importance of β -cell failure in the development and progression of type 2 diabetes. *J. Clin. Endocrinol. Metab.* **86**, 4047–4058 (2001).
- Moran, A. *et al.* Insulin resistance during puberty: results from clamp studies in 357 children. *Diabetes* **48**, 2039–2044 (1999).
- Buchanan, T. A., Metzger, B. E., Freinkel, N. & Bergman, R. N. Insulin sensitivity and B-cell responsiveness to glucose during late pregnancy in lean and moderately obese women with normal glucose tolerance or mild gestational diabetes. *Am. J. Obstet. Gynecol.* **162**, 1008–1014 (1990).
- DeFronzo, R. A. Glucose intolerance of aging. Evidence for tissue insensitivity to insulin. *Diabetes* **28**, 1095–1101 (1979).
- Goodyear, L. J. & Kahn, B. B. Exercise, glucose transport, and insulin sensitivity. *Annu. Rev. Med.* **49**, 235–261 (1998).
- Chen, M., Bergman, R. N. & Porte, D. Insulin resistance and β -cell dysfunction in aging: the importance of dietary carbohydrate. *J. Clin. Endocrinol. Metab.* **67**, 951–957 (1988).
- Wellen, K. E. & Hotamisligil, G. S. Inflammation, stress, and diabetes. *J. Clin. Invest.* **115**, 1111–1119 (2005).
- Scherer, P. E. Adipose tissue: from lipid storage compartment to endocrine organ. *Diabetes* **55**, 1537–1545 (2006).
- Shoelson, S. E., Lee, J. & Goldfine, A. B. Inflammation and insulin resistance. *J. Clin. Invest.* **116**, 1793–1801 (2006).
- Yang, Q. *et al.* Serum retinol binding protein 4 contributes to insulin resistance in obesity and type 2 diabetes. *Nature* **436**, 356–362 (2005).

18. Kadowaki, T. *et al.* Adiponectin and adiponectin receptors in insulin resistance, diabetes, and the metabolic syndrome. *J. Clin. Invest.* **116**, 1784–1792 (2006).
19. Fain, J. N., Madan, A. K., Hiler, M. L., Cheema, P. & Bahouth, S. W. Comparison of the release of adipokines by adipose tissue, adipose tissue matrix, and adipocytes from visceral and subcutaneous abdominal adipose tissues of obese humans. *Endocrinology* **145**, 2273–2282 (2004).
20. Mooney, R. A. *et al.* Suppressors of cytokine signaling-1 and -6 associate with and inhibit the insulin receptor. A potential mechanism for cytokine-mediated insulin resistance. *J. Biol. Chem.* **276**, 25889–25893 (2001).
21. Perreault, M. & Marette, A. Targeted disruption of inducible nitric oxide synthase protects against obesity-linked insulin resistance in muscle. *Nature Med.* **7**, 1138–1143 (2001).
22. Weisberg, S. P. *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *J. Clin. Invest.* **112**, 1796–1808 (2003).
23. Xu, H. *et al.* Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J. Clin. Invest.* **112**, 1821–1830 (2003).
24. Reaven, G. M., Hollenbeck, C., Jeng, C. Y., Wu, M. S. & Chen, Y. D. Measurement of plasma glucose, free fatty acid, lactate, and insulin for 24 h in patients with NIDDM. *Diabetes* **37**, 1020–1024 (1988).
25. Boden, G. Role of fatty acids in the pathogenesis of insulin resistance and NIDDM. *Diabetes* **46**, 3–10 (1997).
26. Roden, M. *et al.* Mechanism of free fatty acid-induced insulin resistance in humans. *J. Clin. Invest.* **97**, 2859–2865 (1996).
27. Santomauro, A. T. *et al.* Overnight lowering of free fatty acids with Acipimox improves insulin resistance and glucose tolerance in obese diabetic and nondiabetic subjects. *Diabetes* **48**, 1836–1841 (1999).
28. Randle, P. J., Garland, P. B., Hales, C. N. & Newsholme, E. A. The glucose fatty-acid cycle: its role in insulin sensitivity and the metabolic disturbances of diabetes mellitus. *Lancet* **i**, 785–789 (1963).
29. Shulman, G. I. Cellular mechanisms of insulin resistance. *J. Clin. Invest.* **106**, 171–176 (2000).
30. Carey, D. G., Jenkins, A. B., Campbell, L. V., Freund, J. & Chisholm, D. J. Abdominal fat and insulin resistance in normal and overweight women: direct measurements reveal a strong relationship in subjects at both low and high risk of NIDDM. *Diabetes* **45**, 633–638 (1996).
31. Cnop, M. *et al.* The concurrent accumulation of intra-abdominal and subcutaneous fat explains the association between insulin resistance and plasma leptin concentrations: distinct metabolic effects of two fat compartments. *Diabetes* **51**, 1005–1015 (2002).
32. Kahn, S. E. The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of type 2 diabetes. *Diabetologia* **46**, 3–19 (2003).
33. Fujimoto, W. Y. *et al.* Preventing diabetes — applying pathophysiological and epidemiological evidence. *Br. J. Nutr.* **84** (Suppl. 2), S173–S176 (2000).
34. Maeda, K. *et al.* Analysis of an expression profile of genes in the human adipose tissue. *Gene* **190**, 227–235 (1997).
35. Motoshima, H. *et al.* Differential regulation of adiponectin secretion from cultured human omental and subcutaneous adipocytes: effects of insulin and rosiglitazone. *J. Clin. Endocrinol. Metab.* **87**, 5662–5667 (2002).
36. Reynisdottir, S., Dazout, M., Thorne, A. & Langin, D. Comparison of hormone-sensitive lipase activity in visceral and subcutaneous human adipose tissue. *J. Clin. Endocrinol. Metab.* **82**, 4162–4166 (1997).
37. Montague, C. T. & O'Rahilly, S. The perils of portliness: causes and consequences of visceral adiposity. *Diabetes* **49**, 883–888 (2000).
38. Kim, S. P., Ellmerer, M., Van Citters, G. W. & Bergman, R. N. Primacy of hepatic insulin resistance in the development of the metabolic syndrome induced by an isocaloric moderate-fat diet in the dog. *Diabetes* **52**, 2453–2460 (2003).
39. Kloppel, G., Lohr, M., Habich, K., Oberholzer, M. & Heitz, P. U. Islet pathology and the pathogenesis of type 1 and type 2 diabetes mellitus revisited. *Surv. Synth. Pathol. Res.* **4**, 110–125 (1985).
40. Butler, A. E. *et al.* β -cell deficit and increased β -cell apoptosis in humans with type 2 diabetes. *Diabetes* **52**, 102–110 (2003).
41. Chen, C., Hosokawa, H., Bumbalo, L. M. & Leahy, J. L. Mechanism of compensatory hyperinsulinemia in normoglycemic insulin-resistant spontaneously hypertensive rats. Augmented enzymatic activity of glucokinase in β -cells. *J. Clin. Invest.* **94**, 399–404 (1994).
42. Liu, Y. Q., Jetton, T. L. & Leahy, J. L. β -cell adaptation to insulin resistance. Increased pyruvate carboxylase and malate-pyruvate shuttle activity in islets of nondiabetic Zucker fatty rats. *J. Biol. Chem.* **277**, 39163–39168 (2002).
43. Kahn, S. E. *et al.* Increased β -cell secretory capacity as mechanism for islet adaptation to nicotinic acid-induced insulin resistance. *Diabetes* **38**, 562–568 (1989).
44. Kahn, S. E. *et al.* Effect of exercise on insulin action, glucose tolerance and insulin secretion in aging. *Am. J. Physiol.* **258**, E937–E943 (1990).
45. Dobbins, R. L. *et al.* A fatty acid-dependent step is critically important for both glucose — and non-glucose — stimulated insulin secretion. *J. Clin. Invest.* **101**, 2370–2376 (1998).
46. Prentki, M., Joly, E., El-Asaad, W. & Roduit, R. Malonyl-CoA signaling, lipid partitioning, and glucolipotoxicity: role in β -cell adaptation and failure in the etiology of diabetes. *Diabetes* **51** (Suppl. 3), S405–S413 (2002).
47. Bergman, R. N. *et al.* Why visceral fat is bad: mechanisms of the metabolic syndrome. *Obesity* **14** (Suppl. 1), 16S–19S (2006).
48. Itoh, Y. *et al.* Free fatty acids regulate insulin secretion from pancreatic β cells through GPR40. *Nature* **422**, 173–176 (2003).
49. Drucker, D. J. The biology of incretin hormones. *Cell Metab.* **3**, 153–165 (2006).
50. Verdich, C. *et al.* The role of postprandial releases of insulin and incretin hormones in meal-induced satiety — effect of obesity and weight reduction. *Int. J. Obes. Relat. Metab. Disord.* **25**, 1206–1214 (2001).
51. Berthoud, H. R. & Jeanrenaud, B. Acute hyperinsulinemia and its reversal by vagotomy after lesions of the ventromedial hypothalamus in anesthetized rats. *Endocrinology* **105**, 146–151 (1979).
52. Ahren, B., Taborsky, G. J. & Porte, D. Neuropeptidergic versus cholinergic and adrenergic regulation of islet hormone secretion. *Diabetologia* **29**, 827–836 (1986).
53. Hull, R. L. *et al.* Dietary-fat-induced obesity in mice results in beta cell hyperplasia but not increased insulin release: evidence for specificity of impaired beta cell adaptation. *Diabetologia* **48**, 1350–1358 (2005).
54. Sorenson, R. L. & Brelje, T. C. Adaptation of islets of Langerhans to pregnancy: beta-cell growth, enhanced insulin secretion and the role of lactogenic hormones. *Horm. Metab. Res.* **29**, 301–307 (1997).
55. Bonner-Weir, S., Deery, D., Leahy, J. L. & Weir, G. C. Compensatory growth of pancreatic β -cells in adult rats after short-term glucose infusion. *Diabetes* **38**, 49–53 (1989).
56. Steil, G. M. *et al.* Adaptation of β -cell mass to substrate oversupply: enhanced function with normal gene expression. *Am. J. Physiol. Endocrinol. Metab.* **280**, E788–E796 (2001).
57. Rhodes, C. J. Type 2 diabetes — a matter of β -cell life and death? *Science* **307**, 380–384 (2005).
58. Bernal-Mizrachi, E., Wen, W., Stahlhut, S., Welling, C. M. & Permutt, M. A. Islet β cell expression of constitutively active Akt1/PKB α induces striking hypertrophy, hyperplasia, and hyperinsulinemia. *J. Clin. Invest.* **108**, 1631–1638 (2001).
59. Hennige, A. M. *et al.* Upregulation of insulin receptor substrate-2 in pancreatic β cells prevents diabetes. *J. Clin. Invest.* **112**, 1521–1532 (2003).
60. Kiba, T. *et al.* Ventromedial hypothalamic lesion-induced vagal hyperactivity stimulates rat pancreatic cell proliferation. *Gastroenterology* **110**, 885–893 (1996).
61. Røder, M. E., Porte, D. & Kahn, S. E. Disproportionately elevated proinsulin levels reflect the degree of impaired B-cell secretory capacity in patients with non-insulin dependent diabetes mellitus. *J. Clin. Endocrinol. Metab.* **83**, 604–608 (1998).
62. Garvey, W. T., Olefsky, J. M., Griffen, J., Hamman, R. F. & Kolterman, O. G. The effect of insulin treatment on insulin secretion and insulin action in type II diabetes mellitus. *Diabetes* **34**, 222–234 (1985).
63. Kahn, S. E., Bergman, R. N., Schwartz, M. W., Taborsky, G. J. & Porte, D. Short-term hyperglycemia and hyperinsulinemia improve insulin action but do not alter glucose action in normal humans. *Am. J. Physiol.* **262**, E518–E523 (1992).
64. Sako, Y. & Grill, V. E. A 48-hour lipid infusion in the rat time-dependently inhibits glucose-induced insulin secretion and B cell oxidation through a process likely coupled to fatty acid oxidation. *Endocrinology* **127**, 1580–1589 (1990).
65. Zhou, Y. P. & Grill, V. E. Long-term exposure of rat pancreatic islets to fatty acids inhibits glucose-induced insulin secretion and biosynthesis through a glucose fatty acid cycle. *J. Clin. Invest.* **93**, 870–876 (1994).
66. Carpentier, A. *et al.* Acute enhancement of insulin secretion by FFA in humans is lost with prolonged FFA elevation. *Am. J. Physiol.* **276**, E1055–E1066 (1999).
67. Utzschneider, K. M. *et al.* Impact of differences in fasting glucose and glucose tolerance on the hyperbolic relationship between insulin sensitivity and insulin responses. *Diabetes Care* **29**, 356–362 (2006).
68. Ward, W. K. *et al.* Insulin resistance and impaired insulin secretion in subjects with histories of gestational diabetes mellitus. *Diabetes* **34**, 861–869 (1985).
69. Ehrmann, D. A. *et al.* Insulin secretory defects in polycystic ovary syndrome. Relationship to insulin sensitivity and family history of non-insulin-dependent diabetes mellitus. *J. Clin. Invest.* **96**, 520–527 (1995).
70. Kahn, S. E. *et al.* Exercise training delineates the importance of B-cell dysfunction to the glucose intolerance of human aging. *J. Clin. Endocrinol. Metab.* **74**, 1336–1342 (1992).
71. Cavaghan, M. K., Ehrmann, D. A., Byrne, M. M. & Polonsky, K. S. Treatment with the oral antidiabetic agent troglitazone improves β cell responses to glucose in subjects with impaired glucose tolerance. *J. Clin. Invest.* **100**, 530–537 (1997).
72. Jensen, C. C. *et al.* β -cell function is the major determinant of oral glucose tolerance in four ethnic groups in the United States. *Diabetes* **51**, 2170–2178 (2002).
73. Knowles, N. G., Landchild, M. A., Fujimoto, W. Y. & Kahn, S. E. Insulin and amylin release are both diminished in first-degree relatives of subjects with type 2 diabetes. *Diabetes Care* **25**, 292–297 (2002).
74. Weyer, C., Bogardus, C., Mott, D. M. & Pratley, R. E. The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. *J. Clin. Invest.* **104**, 787–794 (1999).
75. Festa, A., Williams, K., D'Agostino, R., Wagenknecht, L. E. & Haffner, S. M. The natural course of β -cell function in nondiabetic and diabetic individuals: the Insulin Resistance Atherosclerosis Study. *Diabetes* **55**, 1114–1120 (2006).
76. O'Rahilly, S. & Farooqi, I. S. Genetics of obesity. *Phil. Trans. R. Soc. B* **361**, 1095–1105 (2006).
77. Barroso, I. Genetics of type 2 diabetes. *Diabet. Med.* **22**, 517–535 (2005).
78. Andriulionyte, L., Zacharova, J., Chiasson, J. L. & Laakso, M. Common polymorphisms of the PPAR- γ 2 (Pro12Ala) and PGC-1 α (Gly482Ser) genes are associated with the conversion from impaired glucose tolerance to type 2 diabetes in the STOP-NIDDM trial. *Diabetologia* **47**, 2176–2184 (2004).
79. Kaiyala, K. J. *et al.* Reduced β -cell function contributes to impaired glucose tolerance in dogs made obese by high-fat feeding. *Am. J. Physiol.* **277**, E659–E667 (1999).
80. Hales, C. N. & Barker, D. J. Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Diabetologia* **35**, 595–601 (1992).
81. Leung, N. *et al.* Prolonged increase of plasma non-esterified fatty acids fully abolishes the stimulatory effect of 24 hours of moderate hyperglycaemia on insulin sensitivity and pancreatic beta-cell function in obese men. *Diabetologia* **47**, 204–213 (2004).
82. Schwartz, M. W., Woods, S. C., Porte, D., Seeley, R. J. & Baskin, D. G. Central nervous system control of food intake. *Nature* **404**, 661–671 (2000).
83. Otani, K. *et al.* Reduced β -cell mass and altered glucose sensing impair insulin-secretory function in β IRKO mice. *Am. J. Physiol. Endocrinol. Metab.* **286**, E41–E49 (2004).

Acknowledgements This work was supported in part by the US Department of Veterans Affairs and the NIH. S.E.K. is the recipient of an American Diabetes Association Distinguished Clinical Scientist Award.

Author information Reprints and permissions information is available at ngp.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to S.E.K. (skahn@u.washington.edu).

Adipocytes as regulators of energy balance and glucose homeostasis

Evan D. Rosen¹ & Bruce M. Spiegelman²

Adipocytes have been studied with increasing intensity as a result of the emergence of obesity as a serious public health problem and the realization that adipose tissue serves as an integrator of various physiological pathways. In particular, their role in calorie storage makes adipocytes well suited to the regulation of energy balance. Adipose tissue also serves as a crucial integrator of glucose homeostasis. Knowledge of adipocyte biology is therefore crucial for understanding the pathophysiological basis of obesity and metabolic diseases such as type 2 diabetes. Furthermore, the rational manipulation of adipose physiology is a promising avenue for therapy of these conditions.

Adipose tissue has been ignored by anatomists and physicians for centuries, considered to be an energy storage depot with few interesting attributes. The well-documented rise in obesity during the past 30 years¹ has contributed to the negative image of adipose tissue, particularly in the popular mind. The past two decades, however, have seen a wave of intense scientific interest in this cell type, fuelled in part by concerns about obesity and its attendant metabolic sequelae², and also by the recognition that adipocytes integrate a wide array of homeostatic processes. In addition to regulating fat mass and nutrient homeostasis (discussed below), adipocytes are involved in the immune response, blood pressure control, haemostasis, bone mass, and thyroid and reproductive function³. These processes are coordinated mainly through the synthesis and release of peptide hormones by adipocytes.

Adipocytes also release fatty acids into the circulation, which are used by most organs for fuel when glucose is limiting. These fatty acids are generated by breaking down triacylglycerols, which contain more energy per unit mass than do carbohydrates and can essentially be stored anhydrously. By contrast, glycogen has only the half the energy content per unit of pure mass, and must be stored in association with water, further decreasing its efficiency.

Although most multicellular organisms have cells that store excess energy, adipocytes evolved to meet this need at the time of the vertebrate radiation. Mammals, birds, reptiles, amphibians and many (but not all) fish have cells that are readily identifiable as adipocytes, although the anatomical location of fat tissues varies considerably between species⁴. Most mammals have stereotypical adipose depots located throughout the body. Some of these depots are thought to be largely structural, providing mechanical support but contributing relatively little to energy homeostasis. Examples include the fat pads of the heels, fingers and toes, and the periorbital fat supporting the eyes. Other adipocytes exist in loose association with the skin, and have been termed subcutaneous fat. These cells are the cause of 'cellulite', and are the target of cosmetic procedures such as liposuction. Finally, there are several distinct depots within the body cavity, surrounding the heart and other organs, associated with the intestinal mesentery, and in the retroperitoneum. Some of these depots, known as visceral fat, drain directly into the portal circulation and have been linked to many of the morbidities associated with obesity, including type 2 diabetes and cardiovascular disease.

Adipocytes and precursor cells from different depots have different replicative potential, different developmental attributes and different responses to hormonal signals, although the mechanistic basis for these distinctions is still unclear⁵.

In addition to depot-specific differences, a further distinction must be made between brown and white adipocytes. Brown adipocytes are found only in mammals, and differ from the more typical white adipocytes in that they express uncoupling protein-1 (UCP-1), which dissipates the proton gradient across the inner mitochondrial membrane that is produced by the action of the electron transport chain. This generates heat at the expense of ATP. Morphologically, brown adipocytes are multilocular and contain less overall lipid than their white counterparts, and are particularly rich in mitochondria. Rodents have a distinct brown fat pad, which lies in the interscapular region. In humans, brown adipose tissue surrounds the heart and great vessels in infancy but tends to disappear over time until only scattered cells can be found within white fat pads.

This review briefly examines the transcriptional basis of adipocyte development, and then discusses energy homeostasis in mammals and how adipocytes regulate components of that system. The second part of the review provides a similar look at the role of adipose tissue in glucose homeostasis. Adipocytes have a crucial role in regulating both of these physiological processes through a series of endocrine and non-endocrine mechanisms. These involve a widening array of adipose-derived secreted molecules (known as adipokines), neural connections and changes in whole-body physiology wrought by primary alterations in adipocyte cellular metabolism.

Transcriptional regulation of adipocyte differentiation

Adipocytes have been a popular model for the study of cell differentiation since the development of the murine adipose 3T3 cell culture system by Green and colleagues⁶. There have been several thorough reviews on this aspect of adipose biology recently^{7,8}, so we present only the core of this regulatory system. The central engine of adipose differentiation is peroxisome proliferator-activated receptor- γ (PPAR- γ)^{9–11}. When this receptor is activated by an agonist ligand in fibroblastic cells, a full programme of differentiation is stimulated, including morphological changes, lipid accumulation and the expression of almost all genes characteristic of fat cells. Multiple CCAAT/enhancer-binding proteins (C/EBPs) also

¹Division of Endocrinology and Metabolism, Beth Israel Deaconess Medical Centre, 330 Brookline Avenue, Boston, Massachusetts 02215, USA. ²Dana-Farber Cancer Institute, 1 Jimmy Fund Way, Boston, Massachusetts 02115, USA.

have a critical role in adipogenesis, with C/EBP- β and C/EBP- δ driving PPAR- γ expression in the early stages of differentiation and C/EBP α maintaining PPAR- γ expression later on in the process⁷. C/EBPs and PPAR- γ also directly activate many of the genes of terminally differentiated adipocytes. More recently, other factors have been implicated in the differentiation process, including several Krüppel-like factors^{12–14}, early growth response 2 (Krox20)¹⁵ and early B-cell factors¹⁶.

This transcriptional cascade seems to function in both white and brown adipogenesis. However, ectopic expression of PPAR- γ or C/EBP proteins in fibroblasts results in the formation of adipocytes with the characteristics of white fat cells, with little or no expression of UCP-1 even when stimulated with β -adrenergic agents. Relatively little is known about brown fat differentiation, except for the important role of transcriptional cofactors, including pRb (ref. 17), p107 (ref. 18), SRC-1 and TIF2 (ref. 19), and PPAR- γ -coactivator-1 α (PGC-1 α)²⁰. PGC-1 α coactivates PPAR- γ and other transcription factors, is expressed at much higher levels in brown adipocytes than in white adipocytes, and is induced in brown cells upon cold exposure *in vivo*, or by β -adrenergic stimulation in isolated cells^{20,21}. When introduced into white fat cells in culture or *in vivo*, PGC-1 α 'switches on' many of the key features of brown cells, including mitochondrial biogenesis and UCP-1 activity²⁰. Cells ectopically expressing PGC-1 α also have an increased fraction of uncoupled respiration, a key characteristic of brown fat. Although PGC-1 α is clearly a key effector of the thermogenic programme of brown fat, cells lacking PGC-1 α still show a morphology that is brown-fat-like, and still express certain molecular markers of brown fat²². Thus, it seems likely that other factors function upstream of PGC-1 α to control the determination of brown fat cells.

Principles of energy balance

Energy balance in animals is governed by the First Law of Thermodynamics, and is often expressed as a simple equation:

$$\text{Energy intake} = \text{energy burned} + \text{energy stored} \quad (1)$$

Lipid storage in adipose tissue thus represents excess energy consumption relative to energy expenditure (Fig. 1). Although fundamentally true, this simple representation overlooks a few key features of energy homeostasis *in vivo*. First, although food intake is relatively easy to measure, it is not the precise parameter that determines the amount of energy brought into the system. The efficiency of calorie absorption in the gut, which is much more difficult to measure and is usually ignored in practice, must also be accounted for. A second consideration is that the body's response to alterations in energy input or expenditure is not static. In general, energy homeostasis is regulated to defend the highest weight achieved²³. Thus voluntary reductions in food intake are countered by involuntary reductions in energy expenditure, making weight loss more difficult than a simple interpretation of equation (1) would indicate. Overall, energy balance is responsive to various factors, including hormones and neural inputs, in addition to psychological and cultural factors²⁴.

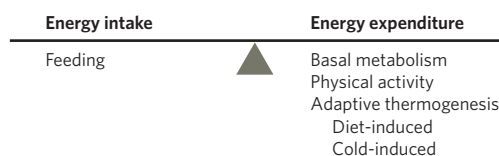


Figure 1 | Energy homeostasis depends upon the balance between caloric intake and energy expenditure. Although caloric intake is almost entirely due to the consumption of food (minus whatever fails to be absorbed), energy expenditure has more components, including basal metabolism, physical activity (voluntary and involuntary) and adaptive thermogenesis. The last category includes the small amount of energy spent in absorbing and processing the diet (known as diet-induced thermogenesis) as well as energy spent to maintain body temperature in the face of cold.

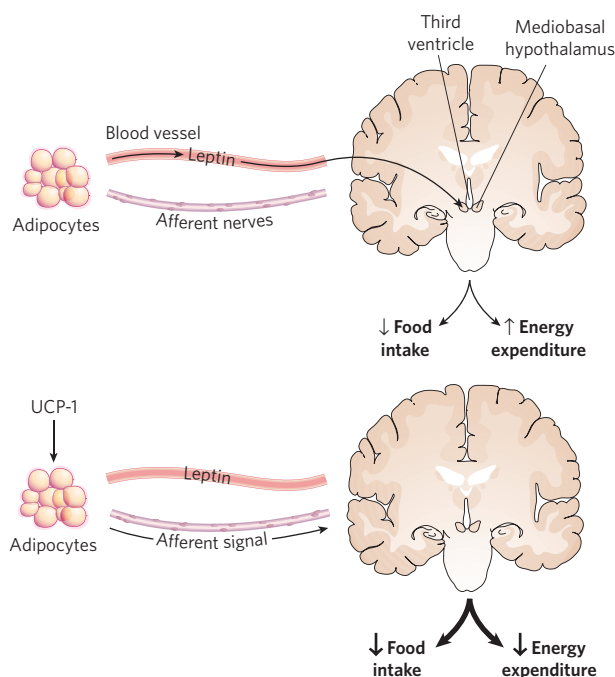


Figure 2 | Adipocytes regulate energy balance by endocrine and non-endocrine mechanisms. Adipocytes synthesize and secrete leptin, which circulates in the blood and acts on the CNS (primarily the hypothalamus) to reduce food intake and enhance energy expenditure. Adipose tissue also communicates with the CNS by means of a rich network of peripheral nerves, which can transmit afferent signals about energy status to the brain, such as when UCP-1 is expressed ectopically. This results in enhanced leptin sensitivity, which increases the effect of secreted leptin on energy balance.

Adipocytes and energy balance

Adipose tissue contains most of the energy stores in normal healthy humans. An important but underappreciated fact is that having more fat cells does not make an animal fatter. In the absence of altered energy balance, an increase in adipogenesis will result in smaller fat cells with no change in total adiposity. Conversely, a reduction in adipocyte number without a change in energy balance would result in larger fat cells, but not less total adipose mass. For example, surgical removal of fat can have cosmetic effects but does not change the energy balance equation. Careful studies in animals have shown that total body fat usually recovers after surgical removal of fat pads²⁵. If certain depots are removed entirely, the fat usually increases at other anatomical locations, though careful clinical studies have not yet been reported in humans. The fact that adipocyte differentiation does not in itself cause obesity does not mean that adipocytes have no role in energy balance. In fact, adipose tissues are critical integrators of organismal energy balance, through the regulation of both food intake and energy expenditure. These effects are mediated by both endocrine and non-endocrine means.

Leptin

Leptin was the first adipokine discovered to have a role in modulating adiposity, and it remains the best studied^{26–29}. Leptin is secreted almost exclusively by fat, and serves as a major 'adipostat' by repressing food intake and promoting energy expenditure. Predictably, animals and humans with mutations in either leptin or the leptin receptor are obese. The leptin receptor is expressed at low levels in numerous tissues, but is found at high levels in the mediobasal hypothalamus, particularly the arcuate nucleus, the ventromedial nucleus and the dorsomedial nucleus^{30–32} (Fig. 2). Leptin receptor activation at these sites leads to repression of orexigenic pathways (for example, those involving neuropeptide Y, NPY, and agouti-related peptide, AgRP) and induction of

anorexigenic pathways (such as those involving pro-opiomelanocortin, POMC, and cocaine and amphetamine-regulated transcript, CART). Leptin-dependent effects on food intake and energy expenditure ultimately diverge in the central nervous system (CNS), partly at the level of melanocortin MC4 receptor signalling³³. Recently, increased attention has been paid to leptin action in non-hypothalamic sites such as the caudal brainstem³⁴.

Neural pathways

Fat pads are richly innervated by sympathetic fibres, and activation of these fibres is associated with enhanced lipolysis^{35,36}. These nerves also regulate fat pad cellularity — denervation of specific depots results in a twofold elevation in adipocyte number in hamsters and rats³⁵. Recent studies have also suggested that neural afferent signals from adipose tissue to the brain might also regulate adiposity. Direct introduction of low levels of UCP-1 into white adipose tissue causes a marked increase in leptin sensitivity in mice, with a concomitant reduction in food intake, and enhanced energy expenditure and weight loss³⁷ (Fig. 2). This effect is abrogated entirely when the manipulated fat pad is denervated. The actual parameter being monitored by the adipocyte is unclear, but might include reactive oxygen species (ROS), ATP levels or even local heat production.

Alterations in adipocyte metabolism

Another way in which adipocytes can modulate whole-body energy balance is through alterations in their own metabolism. This is perhaps best understood for brown fat. Brown adipocytes are highly specialized to perform uncoupled respiration, which dissipates chemical energy in the form of heat. In classical mitochondrial dynamics, fuel oxidation is linked to electron transport, resulting in the development of an electrochemical gradient across the inner mitochondrial membrane. This gradient, which is formed by the extrusion of protons across the inner mitochondrial membrane by three protein complexes in the electron transport system, is normally dissipated at complex V, which couples proton inflow to ATP synthesis. Brown fat, however, expresses UCP-1, which allows protons to flow back across the inner mitochondrial membrane without concomitant ATP synthesis, resulting in heat generation. In rodents, brown adipose tissue makes a substantial contribution to whole-body energy metabolism. Mice lacking brown adipose tissue exhibit reduced energy expenditure and are prone to diet-induced obesity^{38,39}. Interestingly, mice lacking UCP-1 are cold-sensitive but not obese⁴⁰. Taken together, these genetic models suggest that brown adipose tissue might influence whole-body energy metabolism in ways not fully explained by the expression of UCP-1. It is clear, however, that overexpression of UCP-1 in white adipose tissue results in reduced adiposity. As mentioned earlier, at least part of this effect is neurally mediated³⁷.

Enhancement of fatty acid oxidation in white adipocytes (without uncoupling) has also been proposed to reduce adiposity. Early examples of genetically altered mice with enhanced adipocyte fatty-acid oxidation and concomitant leanness were identified in global knockouts, making it difficult to pinpoint altered metabolism in adipose tissue as the primary cause of the leanness⁴¹. In one example, ablation of the transcriptional co-repressor receptor-interacting protein 140 (RIP140) results in lean mice with enhanced fatty-acid oxidation and oxygen consumption associated with increased UCP-1 expression in white adipocytes⁴². Another recent study showed that adipose-specific targeting of the pseudokinase TRB3 results in enhanced lipolysis and fatty-acid oxidation, with subsequent protection from diet-induced obesity⁴³. This effect is caused by the TRB3-mediated ubiquitination and degradation of acetyl-coenzyme-A carboxylase (ACC), which catalyses the rate-limiting step in lipogenesis. Enhanced lipolysis alone is insufficient to promote weight loss; the liberated fatty acids must be oxidized. Furthermore, oxidation alone is also insufficient unless the ATP generated is consumed. Thus, the observed weight loss in the *Trb3* transgenic mice leads us to infer that there is undetected uncoupling in the mitochondria or that an ATP-consuming process has been activated.

Principles of glucose homeostasis

Despite intermittent ingestion of dietary carbohydrates, serum glucose levels remain relatively steady throughout the day. This requires the concerted actions of several different tissues^{44,45} (Fig. 3). Pancreatic β -cells, for example, secrete insulin in response to the elevations in glucose that occur after eating. Insulin promotes glucose disposal in adipose tissue and muscle, and also prevents the liver from producing more glucose by suppressing glycogenolysis and gluconeogenesis. In the fasting state, low insulin levels combined with elevated counter-regulatory hormones such as glucagon, adrenaline and corticosteroids promote hepatic glucose production. Recently, evidence has emerged that the brain coordinates many of these effects as well, through direct and indirect glucose sensing and neural outputs to peripheral organs⁴⁴.

Adipocytes as regulators of glucose homeostasis

At first, the idea that adipose tissue could have a considerable effect on global glycaemic control was not easy to accept. Early studies determined that adipose tissue accounts for only a fraction of glucose disposal after a meal (about 10–15%), with most of the rest taken up by muscle⁴⁶. Nonetheless, it was equally clear that alterations in adiposity have profound implications for glucose homeostasis; too much fat (obesity) and too little fat (lipodystrophy) are both associated with insulin resistance and hyperglycaemia. In addition, PPAR- γ ligands such as thiazolidinedione (TZD) drugs have excellent anti-diabetic activity despite the fact that most PPAR- γ is found in adipose tissue and not muscle. We now understand that the profound effect of adipocytes on glucose balance is mediated by several different mechanisms, which can be loosely categorized (as before with energy balance) as endocrine (Fig. 4) and non-endocrine.

Leptin

As described above, leptin is a multifunctional protein secreted primarily by adipocytes. In addition to its well-described role in energy balance, leptin has notable effects on glucose homeostasis, as it reverses

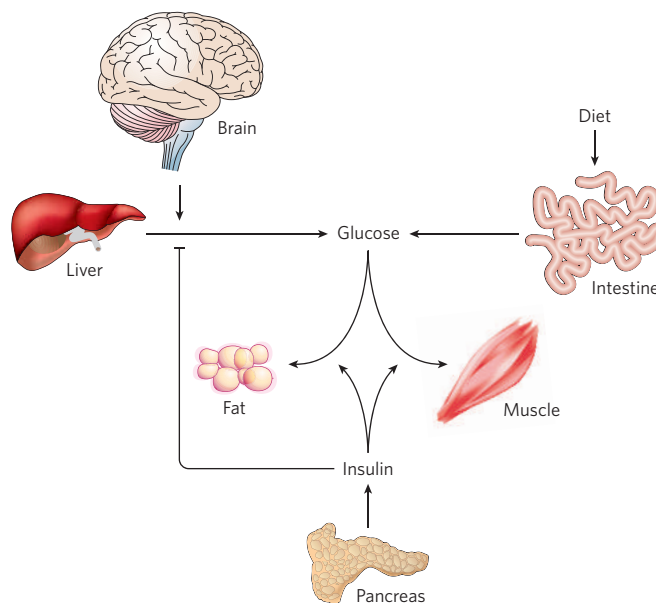


Figure 3 | Glucose homeostasis requires the coordinated actions of various organs. Inputs to serum glucose levels include absorption from the intestine and release from the liver. The latter occurs by breakdown of preformed glycogen as well as gluconeogenesis, and both processes are inhibited by insulin. Glucose is removed from the system by uptake into virtually all cell types, but most importantly into muscle and adipose tissue, which requires insulin. Recent evidence suggests that the CNS can also sense glucose and act to affect systemic glycaemia, at least in part by regulating gluconeogenesis.

hyperglycaemia in *ob/ob* mice before body weight is corrected²⁹. Similarly, pair feeding *ob/ob* mice to match control animals does not restore glucose tolerance as well as exogenous leptin does⁴⁷. Leptin also improves glucose homeostasis in lipodystrophic mice, and in humans with lipodystrophy or congenital leptin deficiency^{48–50}. Results in humans with 'typical' obesity have been disappointing in this regard⁵¹. However, this is consistent with the high levels of leptin and apparent leptin resistance seen in these patients.

The anti-hyperglycaemic actions of leptin are mediated through several different organs. First, leptin improves insulin sensitivity in muscles. Leptin also reduces intra-myocellular lipid levels through a combination of direct activation of AMP-activated protein kinase (AMPK) and indirect actions mediated through central neural pathways⁵². The effect on lipid partitioning might help to explain the improved insulin sensitivity, according to the current idea that intracellular lipids contribute to insulin resistance. Leptin also improves insulin sensitivity in the liver, an effect seen with either peripheral or intracerebroventricular administration⁵³. As in muscles, leptin functions in part to reduce hepatic intracellular triacylglycerol levels. There has also been discussion of an 'adipo-insular axis', with insulin promoting leptin secretion and leptin inhibiting insulin release⁵⁴. Consistent with this idea, ablation of leptin receptors from β -cells results in enhanced basal insulin secretion and fasting hypoglycaemia⁵⁵. The effect of leptin on insulin levels is due to inhibition of proinsulin synthesis as well as reduction of secretion⁵⁴.

Adiponectin

The hormone adiponectin was identified by several different groups and given various names (for example, apM1, GBP28, AdipoQ and ACRP30)⁵⁶. This 30-kDa protein has an amino-terminal collagen-like domain and a carboxy-terminal globular domain that mediates multimerization. Circulating adiponectin can exist as a trimer, hexamer or a higher-order multimer with 12–18 subunits^{57,58}; there is some controversy about which is the active form of the hormone. Two types of receptor have been proposed. One comprises two similar transmembrane proteins with homology to G-protein-coupled receptors, known as adipoR1 and adipoR2 (ref. 59); a second molecule (T-cadherin) without a transmembrane domain has been proposed to act as a co-receptor for the high-molecular-weight form of adiponectin on endothelial and smooth muscle cells⁶⁰. Interestingly, adiponectin circulates at extraordinarily high concentrations (5–10 $\mu\text{g ml}^{-1}$), accounting for 0.01% of all plasma protein⁶¹. Unlike almost all other adipokines, however, adiponectin levels are inversely correlated with body mass^{62,63}, for reasons that remain obscure.

Delivery of adiponectin to obese, diabetic mice stimulates AMP kinase activity in the liver and skeletal muscle, with profound effects on fatty acid oxidation and insulin sensitivity. Loss-of-function models of adiponectin have been more variable, with different groups reporting reduced insulin sensitivity on both chow and high-fat diets, high-fat diet only, or no effect on insulin action at all^{64–66}. The reason for these differences is unknown.

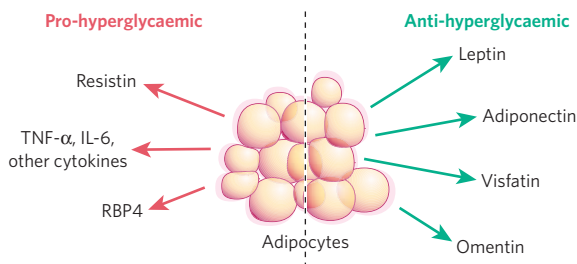


Figure 4 | Adipocytes secrete proteins with varied effects on glucose homeostasis. Adipocyte-derived proteins with anti-diabetic action (green arrows) include leptin, adiponectin, omentin and visfatin. Other factors tend to raise blood glucose (red arrows), including resistin, TNF- α and RBP4. TNF- α and human resistin are probably secreted by non-adipocytes within the fat pad. IL, interleukin.

Two groups have shown that part of the anti-diabetic action of TZDs requires adiponectin^{67,68}; TZDs significantly raise plasma adiponectin levels through effects on synthesis and secretion⁶⁹. Adiponectin has no effect on insulin secretion in islets from healthy mice or humans, but does enhance glucose-stimulated insulin secretion in islets from mice with diet-induced obesity⁷⁰.

Visfatin

The connection between visceral adiposity and insulin resistance has led several groups to try to identify secreted products derived specifically from this depot. The first such protein reported was visfatin, which had been identified in immune cells years earlier as pre-B-cell colony-enhancing factor (PBEF)⁷¹. There were several surprising aspects surrounding this discovery, including the fact that visfatin does not promote insulin resistance — on the contrary, it has a salutary effect on glucose uptake mediated by direct binding and activation of the insulin receptor. Visfatin circulates at concentrations well below those of insulin (<10%), however, and fasting and feeding do not regulate its expression, making it doubtful that visfatin alone is an important factor in insulin-receptor signalling. Other aspects of visfatin biology require further study. For example, serum levels of visfatin are variably correlated with type 2 diabetes and other insulin-resistant states⁷². Visfatin also has no signal sequence and has been shown to have enzymatic activity as a nicotinamide phosphoribosyltransferase with residence in the nucleus and cytosol⁷³. It is not clear whether there is regulated secretion of visfatin or whether serum levels reflect leakage from dead or damaged cells.

Omentin

Omentin is another peptide secreted predominantly by visceral fat. Like visfatin, it has positive effects on glucose uptake, although omentin works as an insulin sensitizer and does not have insulin-mimetic properties⁷⁴. Unlike visfatin, omentin seems to be made by stromal-vascular cells within the fat pad rather than adipocytes. Interestingly, omentin is produced in considerable quantities by adipose tissue in humans and macaques but not mice⁷⁴. Omentin's mechanism of action, including target tissues, a receptor or relevant signal transduction pathways, remains obscure.

Tumour necrosis factor- α and other cytokines

Tumour necrosis factor- α (TNF- α) was the first secreted adipose protein to be shown to have effects on glucose homeostasis⁷⁵. TNF- α levels are elevated in obesity and in other insulin-resistant states (such as sepsis), addition of TNF- α to cells and mice reduces insulin action, and blockade of TNF- α action by biochemical or genetic means restores insulin sensitivity *in vivo* and *in vitro*. Interestingly, TNF- α seems to be derived from cell types other than adipocytes themselves. Macrophages in particular have been implicated in TNF- α production from murine fat pads⁷⁶. Other cytokines, including interleukin-6, are produced by adipocytes, and there is conflicting evidence suggesting that they have both insulin-resistance-promoting and insulin-sensitizing effects^{77,78}. Such 'adipocytokines' can promote insulin resistance through several mechanisms. These include c-Jun N-terminal kinase 1 (JNK1)-mediated serine phosphorylation of insulin receptor substrate-1 (IRS-1)⁷⁹, I κ B kinase (IKK)-mediated nuclear factor- κ B (NF- κ B) activation⁸⁰, induction of suppressor of cytokine signalling 3 (SOCS3)⁸¹ and production of ROS⁸².

Resistin

Resistin is another small inflammatory molecule with hyperglycaemic action. Resistin (also known as FIZZ3) is one of a family of cysteine-rich resistin-like molecules (RELMS)⁸³. Resistin was discovered as a secreted product of mouse adipocytes that was repressed by TZDs⁸⁴. Levels of resistin are elevated in many murine models of obesity, and gain- and loss-of-function studies in rodents support a role for resistin in increasing hepatic glucose output^{85,86}. Resistin might also be involved in reducing glucose uptake by muscles and fat, but this is not as robust an effect as that seen in the liver. As with adiponectin, there are several

multimeric forms of resistin circulating in plasma, and action at the cellular level seems to depend on species with lower molecular weight⁸⁷. There is considerable controversy surrounding the role of resistin in humans, in whom even the cellular source of resistin has been debated. The bulk of the data suggest that human resistin is the product of macrophages or other stromal cells within the fat pad^{88,89}.

Retinol-binding protein 4

Specific ablation of glucose transporter 4 (*Glut4*) in mouse adipose tissue was shown to significantly reduce whole-body insulin sensitivity, whereas transgenic overexpression of *Glut4* had the opposite effect. Because the magnitude of the effect was greater than that predicted by altered glucose flux into adipose tissue alone, an endocrine effect was postulated⁹⁰. Transcriptional profiling of samples from these mice led to the discovery that a secreted member of the lipocalin superfamily, retinol-binding protein 4 (RBP4), was coordinately regulated by the changes in adipocyte *GLUT4* levels. Subsequent studies showed that overexpression of RBP4 impairs hepatic and muscle insulin action, and *Rbp4*^{-/-} mice show enhanced insulin sensitivity⁹⁰. Fenretinide, a drug that enhances RBP4 clearance, also increases insulin sensitivity in mice fed a high-fat diet. Furthermore, high serum RBP4 levels are associated with insulin resistance in obese humans and those with type 2 diabetes as well as in lean, nondiabetic people with a family history of diabetes⁹¹. Much remains to be learned about how RBP4 impairs insulin action, and at present it is not clear whether the process involves the retinoid ligand or some other mechanism.

Non-esterified fatty acids

Although the concept of adipokines is relatively new, non-esterified fatty acids (NEFAs) have long been known to be an adipocyte-derived secreted product. NEFAs are primarily released during fasting as a nutrient source for the rest of the body, and have several actions on glucose homeostasis (Fig. 5). Circulating NEFAs reduce adipocyte and muscle glucose uptake, and also promote hepatic glucose output^{92,93}. The net effect of these actions is to promote lipid burning as a fuel source in most tissues, while sparing carbohydrate for neurons (and red blood cells), which depend on glucose as an energy source. Several mechanisms have been proposed to account for the effects of NEFAs on muscle, liver and adipose tissue, including protein kinase C (PKC) activation^{94,95}, oxidative stress⁹⁶, ceramide formation⁹⁷ and activation of the innate immune system^{98,99}. These pathways are not mutually exclusive, and probably function interdependently.

Because lipolysis in adipocytes is repressed by insulin, insulin resistance from any cause can lead to NEFA elevation, which, in turn, induces additional insulin resistance as part of a vicious cycle. β -cells are also affected by NEFAs, depending in part on the duration of exposure; acutely, NEFAs induce insulin secretion (as after a meal), whereas chronic exposure to NEFAs causes a decrease in insulin secretion¹⁰⁰. This effect is mediated by several mechanisms, including lipotoxicity-induced apoptosis of islet cells¹⁰¹. In β -cells that escape apoptosis, NEFAs decrease glucose sensing by inducing expression of uncoupling protein-2 (UCP-2), which decreases mitochondrial membrane potential, ATP synthesis and insulin secretion¹⁰¹.

Lipid sink

Another way, albeit indirect, in which adipose tissue can affect global glucose homeostasis is by serving as a lipid sink. This idea has emerged from the observation that animals and humans with lipodystrophy (in which adipose tissue fails to develop properly) have ectopic lipid deposition in the liver, muscles, β -cells and other organs, which can lead to insulin resistance and decreased insulin secretion¹⁰². The ability to store large amounts of esterified lipid in a manner that is not toxic to the cell or the organism as a whole might be one of the most critical functions of the adipocyte. However, the ability of exogenous leptin^{49,50}, alone or in combination with adiponectin¹⁰³, to improve whole-body glucose homeostasis in lipodystrophy suggests that the endocrine function of adipose tissue is also crucial to the regulation of glucose homeostasis.

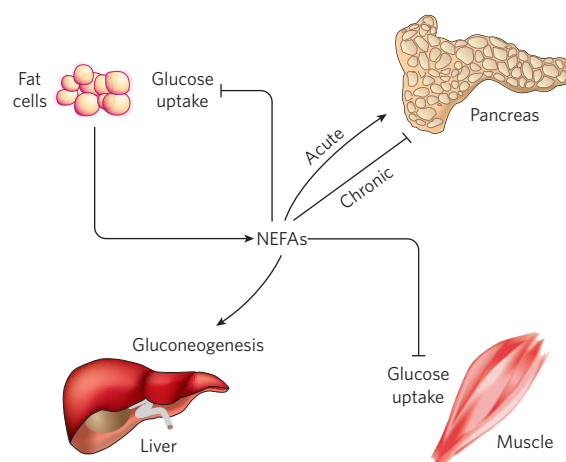


Figure 5 | Adipocyte-derived non-esterified fatty acids have several effects on glucose homeostasis. Lipolysis in adipocytes is repressed by insulin, so it is normal in the fasted state when insulin levels are low. Insulin resistance, however, is also associated with lipolysis and NEFA release into the circulation. Elevated serum NEFAs inhibit insulin's ability to promote peripheral glucose uptake into muscle and fat and to reduce hepatic glucose production. Transiently elevated NEFAs (such as occur after a meal) tend to enhance insulin secretion, whereas chronic elevations in NEFAs (such as occur in insulin resistance) tend to reduce insulin secretion.

Inflammation and adipose tissue in obesity

There is now overwhelming evidence that obesity and type 2 diabetes are inflammatory states, consistent with the production of TNF- α and other cytokines by adipose tissue, as described above. In obese animals and humans, bone-marrow-derived macrophages are recruited to the fat pad under the influence of proteins secreted by adipocytes, including macrophage chemoattractant protein-1 (MCP-1)^{76,104}. Targeted ablation of either MCP-1 or its receptor reduces macrophage infiltration of fat depots and improves insulin sensitivity despite causing no change in body weight¹⁰⁵, and overexpression of MCP-1 has the opposite effect^{106,107}. These data, combined with studies (mentioned above) that indicate a role for pro-inflammatory cytokines in insulin resistance, strongly suggest that intra-adipose macrophages are dominant contributors to the insulin resistance that is associated with obesity. Thiazolidinediones, which are PPAR- γ agonists used clinically as insulin sensitizers, reduce MCP-1 levels and macrophage infiltration into adipose tissue¹⁰⁸.

Alterations in adipocyte metabolism

Genetic manipulations of adipocyte metabolism can have consequences for global glucose homeostasis. Many of these manipulations affect fat mass, which accounts for the observed changes in insulin sensitivity. But the results of some experiments cannot be explained by an effect on adiposity. For example, transgenic overexpression of *Glut4* in adipocytes improves insulin sensitivity (despite a slight increase in body weight)¹⁰⁹, whereas adipose-specific *Glut4*-knockout mice are insulin resistant¹¹⁰. This effect is at least partly explained by changes in RBP4 (see above), although it is not clear whether RBP4 accounts for the full effect. Similarly, the direct injection of UCP-1 into limited areas of white adipose tissue is sufficient to improve whole-body glucose tolerance, an effect that is not neurally mediated and is not dependent on leptin signalling³⁷. An endocrine mechanism is proposed here, but adiponectin levels seem unchanged in these mice, and, so far, a specific effector adipokine has not been discovered.

Future directions

The past decade has seen adipose tissue move from being a minor participant to having a central role in diverse homeostatic processes, with particular emphasis on energy balance and glucose homeostasis. The

plethora of ways in which adipose tissue controls body weight and glycaemia has led to the idea that manipulation of adipocyte biology might be a useful therapeutic strategy in metabolic diseases such as obesity and type 2 diabetes. However, despite the clear connection between excess adiposity and a constellation of health problems, including insulin resistance, dyslipidaemia and cardiovascular disease, it is clear that simply reducing fat cell number is not a viable strategy to promote health. Surgical removal of adipose tissue does not improve metabolic parameters in obese animals or humans. Furthermore, subjects with lipodystrophy are affected by insulin resistance, hyperlipidaemia and steatohepatitis (often leading to cirrhosis) associated with ectopic lipid deposition. A much sounder strategy would be to manipulate adipocyte biology in ways that promote health: enhancement of leptin and adiponectin synthesis, and secretion and promotion of intra-adipocyte lipid oxidation are examples of approaches that might be expected to benefit patients.

The contribution of brown adipose tissue to the total energy balance of humans is largely unexplored. Although infants have recognizable brown adipose tissue depots in the interscapular and around the great vessels, these disappear as humans mature. Individuals with chronically elevated catecholamine secretion, such as those with pheochromocytoma, redevelop distinct brown adipose tissue depots¹¹, as do individuals chronically exposed to cold temperatures. Reactivation of brown adipose tissue in adult humans (or promotion of a brown adipose tissue-like phenotype in white adipocytes) remains a potentially viable therapeutic strategy in humans.

The pace of discovery indicates that this field is still growing rapidly, and new insights are unfolding with each passing month. It remains to be seen how well we will translate these discoveries to our patients, but hopes for breakthrough therapies for obesity and diabetes are as high as they have ever been. ■

- Mokdad, A. H. *et al.* Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. *J. Am. Med. Assoc.* **289**, 76–79 (2003).
- Bray, G. A. & Bellanger, T. Epidemiology, trends, and morbidities of obesity and the metabolic syndrome. *Endocrine* **29**, 109–117 (2006).
- Trayhurn, P. Endocrine and signalling role of adipose tissue: new perspectives on fat. *Acta Physiol. Scand.* **184**, 285–293 (2005).
- Pond, C. M. *The Fats of Life* (Cambridge Univ. Press, Cambridge, 1998).
- Giorgino, F., Laviola, L. & Eriksson, J. W. Regional differences of insulin action in adipose tissue: insights from *in vivo* and *in vitro* studies. *Acta Physiol. Scand.* **183**, 13–30 (2005).
- Green, H. & Kehinde, O. An established preadipocyte cell line and its differentiation in culture. II. Factors affecting the adipose conversion. *Cell* **5**, 19–27 (1975).
- Farmer, S. R. Transcriptional control of adipocyte formation. *Cell Metab.* **4**, 263–273 (2006).
- Hansen, J. B. & Kristiansen, K. Regulatory circuits controlling white versus brown adipocyte differentiation. *Biochem. J.* **398**, 153–168 (2006).
- Rosen, E. D. *et al.* C/EBP α induces adipogenesis through PPAR γ : a unified pathway. *Genes Dev.* **16**, 22–26 (2002).
- Rosen, E. D. *et al.* PPAR γ is required for the differentiation of adipose tissue *in vivo* and *in vitro*. *Mol. Cell* **4**, 611–617 (1999).
- Tontonoz, P., Hu, E. & Spiegelman, B. M. Stimulation of adipogenesis in fibroblasts by PPAR γ , a lipid-activated transcription factor. *Cell* **79**, 1147–1156 (1994).
- Oishi, Y. *et al.* Kruppel-like transcription factor KLF5 is a key regulator of adipocyte differentiation. *Cell Metab.* **1**, 27–39 (2005).
- Mori, T. *et al.* Role of Kruppel-like factor 15 (KLF15) in transcriptional regulation of adipogenesis. *J. Biol. Chem.* **280**, 12867–12875 (2005).
- Banerjee, S. S. *et al.* The Kruppel-like factor KLF2 inhibits peroxisome proliferator-activated receptor- γ expression and adipogenesis. *J. Biol. Chem.* **278**, 2581–2584 (2003).
- Chen, Z., Torrens, J. I., Anand, A., Spiegelman, B. M. & Friedman, J. M. Krox20 stimulates adipogenesis via C/EBP β -dependent and -independent mechanisms. *Cell Metab.* **1**, 93–106 (2005).
- Akerblad, P., Lind, U., Liberg, D., Bamberg, K. & Sigvardsson, M. Early B-cell factor (O/E-1) is a promoter of adipogenesis and involved in control of genes important for terminal adipocyte differentiation. *Mol. Cell Biol.* **22**, 8015–8025 (2002).
- Hansen, J. B. *et al.* Retinoblastoma protein functions as a molecular switch determining white versus brown adipocyte differentiation. *Proc. Natl Acad. Sci. USA* **101**, 4112–4117 (2004).
- Scime, A. *et al.* Rb and p107 regulate preadipocyte differentiation into white versus brown fat through repression of PGC-1 α . *Cell Metab.* **2**, 283–295 (2005).
- Picard, F. *et al.* SRC-1 and TIF2 control energy balance between white and brown adipose tissues. *Cell* **111**, 931–941 (2002).
- Puigserver, P. *et al.* A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell* **92**, 829–839 (1998).
- Puigserver, P. Tissue-specific regulation of metabolic pathways through the transcriptional coactivator PGC1- α . *Int. J. Obes. (Lond.)* **29** (Suppl. 1), S5–S9 (2005).
- Uldry, M. *et al.* Complementary action of the PGC-1 coactivators in mitochondrial biogenesis and brown fat differentiation. *Cell Metab.* **3**, 333–341 (2006).
- Schwartz, M. W. *et al.* Is the energy homeostasis system inherently biased toward weight gain? *Diabetes* **52**, 232–238 (2003).
- Abizaid, A., Gao, Q. & Horvath, T. L. Thoughts for food: brain mechanisms and peripheral energy balance. *Neuron* **51**, 691–702 (2006).
- Mauer, M. M., Harris, R. B. & Bartness, T. J. The regulation of total body fat: lessons learned from lipectomy studies. *Neurosci. Biobehav. Rev.* **25**, 15–28 (2001).
- Friedman, J. M. Leptin and the regulation of body weight. *Harvey Lect.* **95**, 107–136 (1999).
- Friedman, J. M. The function of leptin in nutrition, weight, and physiology. *Nutr. Rev.* **60**, S1–S14 (2002).
- Halaas, J. L. *et al.* Weight-reducing effects of the plasma protein encoded by the obese gene. *Science* **269**, 543–546 (1995).
- Pelleymounter, M. A. *et al.* Effects of the obese gene product on body weight regulation in *ob/ob* mice. *Science* **269**, 540–543 (1995).
- Schwartz, M. W., Seeley, R. J., Campfield, L. A., Burn, P. & Baskin, D. G. Identification of targets of leptin action in rat hypothalamus. *J. Clin. Invest.* **98**, 1101–1106 (1996).
- Bjorbaek, C. & Kahn, B. B. Leptin signaling in the central nervous system and the periphery. *Recent Prog. Horm. Res.* **59**, 305–331 (2004).
- Fei, H. *et al.* Anatomic localization of alternatively spliced leptin receptors (Ob-R) in mouse brain and other tissues. *Proc. Natl Acad. Sci. USA* **94**, 7001–7005 (1997).
- Balthasar, N. *et al.* Divergence of melanocortin pathways in the control of food intake and energy expenditure. *Cell* **123**, 493–505 (2005).
- Grill, H. J. Distributed neural control of energy balance: contributions from hindbrain and hypothalamus. *Obesity (Silver Spring)* **14** (Suppl. 5), 216S–221S (2006).
- Bartness, T. J. & Bamshad, M. Innervation of mammalian white adipose tissue: implications for the regulation of total body fat. *Am. J. Physiol.* **275**, R1399–R1411 (1998).
- Bartness, T. J., Kay Song, C., Shi, H., Bowers, R. R. & Foster, M. T. Brain-adipose tissue cross talk. *Proc. Nutr. Soc.* **64**, 53–64 (2005).
- Yamada, T. *et al.* Signals from intra-abdominal fat modulate insulin and leptin sensitivity through different mechanisms: neuronal involvement in food-intake regulation. *Cell Metab.* **3**, 223–229 (2006).
- Bachman, E. S. *et al.* β AR signaling required for diet-induced thermogenesis and obesity resistance. *Science* **297**, 843–845 (2002).
- Lowell, B. B. *et al.* Development of obesity in transgenic mice after genetic ablation of brown adipose tissue. *Nature* **366**, 740–742 (1993).
- Enerback, S. *et al.* Mice lacking mitochondrial uncoupling protein are cold-sensitive but not obese. *Nature* **387**, 90–94 (1997).
- Abu-Elheiga, L., Matzuk, M. M., Abo-Hashema, K. A. & Wakil, S. J. Continuous fatty acid oxidation and reduced fat storage in mice lacking acetyl-CoA carboxylase 2. *Science* **291**, 2613–2616 (2001).
- Parker, M. G., Christian, M. & White, R. The nuclear receptor co-repressor RIP140 controls the expression of metabolic gene networks. *Biochem. Soc. Trans.* **34**, 1103–1106 (2006).
- Qi, L. *et al.* TRB3 links the E3 ubiquitin ligase COP1 to lipid metabolism. *Science* **312**, 1763–1766 (2006).
- Herman, M. A. & Kahn, B. B. Glucose transport and sensing in the maintenance of glucose homeostasis and metabolic harmony. *J. Clin. Invest.* **116**, 1767–1775 (2006).
- Tirone, T. A. & Brunicaudi, F. C. Overview of glucose regulation. *World J. Surg.* **25**, 461–467 (2001).
- Kahn, B. B. Lilly lecture 1995. Glucose transport: pivotal step in insulin action. *Diabetes* **45**, 1644–1654 (1996).
- Schwartz, M. W. *et al.* Specificity of leptin action on elevated blood glucose levels and hypothalamic neuropeptide Y gene expression in *ob/ob* mice. *Diabetes* **45**, 531–535 (1996).
- Farooqi, I. S. *et al.* Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency. *J. Clin. Invest.* **110**, 1093–1103 (2002).
- Shimomura, I., Hammer, R. E., Ikemoto, S., Brown, M. S. & Goldstein, J. L. Leptin reverses insulin resistance and diabetes mellitus in mice with congenital lipodystrophy. *Nature* **401**, 73–76 (1999).
- Oral, E. A. *et al.* Leptin-replacement therapy for lipodystrophy. *N. Engl. J. Med.* **346**, 570–578 (2002).
- Heymsfield, S. B. *et al.* Recombinant leptin for weight loss in obese and lean adults: a randomized, controlled, dose-escalation trial. *J. Am. Med. Assoc.* **282**, 1568–1575 (1999).
- Minokoshi, Y. *et al.* Leptin stimulates fatty-acid oxidation by activating AMP-activated protein kinase. *Nature* **415**, 339–343 (2002).
- Kamohara, S., Burcelin, R., Halaas, J. L., Friedman, J. M. & Charron, M. J. Acute stimulation of glucose metabolism in mice by leptin treatment. *Nature* **389**, 374–377 (1997).
- Kieffer, T. J. & Habener, J. F. The adipoinsular axis: effects of leptin on pancreatic β -cells. *Am. J. Physiol. Endocrinol. Metab.* **278**, E1–E14 (2000).
- Covey, S. D. *et al.* The pancreatic β cell is a key site for mediating the effects of leptin on glucose homeostasis. *Cell Metab.* **4**, 291–302 (2006).
- Kadowaki, T. *et al.* Adiponectin and adiponectin receptors in insulin resistance, diabetes, and the metabolic syndrome. *J. Clin. Invest.* **116**, 1784–1792 (2006).
- Waki, H. *et al.* Impaired multimerization of human adiponectin mutants associated with diabetes. Molecular structure and multimer formation of adiponectin. *J. Biol. Chem.* **278**, 40352–40363 (2003).
- Pajvani, U. B. *et al.* Structure-function studies of the adipocyte-secreted hormone Acrp30/adiponectin. Implications for metabolic regulation and bioactivity. *J. Biol. Chem.* **278**, 9073–9085 (2003).
- Yamauchi, T. *et al.* Cloning of adiponectin receptors that mediate antidiabetic metabolic effects. *Nature* **423**, 762–769 (2003).
- Hug, C. *et al.* T-cadherin is a receptor for hexameric and high-molecular-weight forms of Acrp30/adiponectin. *Proc. Natl Acad. Sci. USA* **101**, 10308–10313 (2004).
- Arita, Y. *et al.* Paradoxical decrease of an adipose-specific protein, adiponectin, in obesity. *Biochem. Biophys. Res. Commun.* **257**, 79–83 (1999).
- Yatagai, T. *et al.* Hypoadiponectinemia is associated with visceral fat accumulation and insulin resistance in Japanese men with type 2 diabetes mellitus. *Metabolism* **52**, 1274–1278 (2003).
- Hotta, K. *et al.* Circulating concentrations of the adipocyte protein adiponectin are decreased in parallel with reduced insulin sensitivity during the progression to type 2 diabetes in rhesus monkeys. *Diabetes* **50**, 1126–1133 (2001).

64. Kubota, N. *et al.* Disruption of adiponectin causes insulin resistance and neointimal formation. *J. Biol. Chem.* **277**, 25863–25866 (2002).
65. Ma, K. *et al.* Increased β -oxidation but no insulin resistance or glucose intolerance in mice lacking adiponectin. *J. Biol. Chem.* **277**, 34658–34661 (2002).
66. Maeda, N. *et al.* Diet-induced insulin resistance in mice lacking adiponectin/ACRP30. *Nature Med.* **8**, 731–737 (2002).
67. Nawrocki, A. R. *et al.* Mice lacking adiponectin show decreased hepatic insulin sensitivity and reduced responsiveness to peroxisome proliferator-activated receptor γ agonists. *J. Biol. Chem.* **281**, 2654–2660 (2006).
68. Kubota, N. *et al.* Pioglitazone ameliorates insulin resistance and diabetes by both adiponectin-dependent and -independent pathways. *J. Biol. Chem.* **281**, 8748–8755 (2006).
69. Maeda, N. *et al.* PPAR γ ligands increase expression and plasma concentrations of adiponectin, an adipose-derived protein. *Diabetes* **50**, 2094–2099 (2001).
70. Winzell, M. S., Nogueiras, R., Dieguez, C. & Ahren, B. Dual action of adiponectin on insulin secretion in insulin-resistant mice. *Biochem. Biophys. Res. Commun.* **321**, 154–160 (2004).
71. Fukuhara, A. *et al.* Visfatin: a protein secreted by visceral fat that mimics the effects of insulin. *Science* **307**, 426–430 (2005).
72. Stephens, J. M. & Vidal-Puig, A. J. An update on visfatin/pre-B cell colony-enhancing factor, an ubiquitously expressed, illusive cytokine that is regulated in obesity. *Curr. Opin. Lipidol.* **17**, 128–131 (2006).
73. Yang, H., Lavu, S. & Sinclair, D. A. Nampt/PBEF/Visfatin: a regulator of mammalian health and longevity? *Exp. Gerontol.* **41**, 718–726 (2006).
74. Yang, R. Z. *et al.* Identification of omentin as a novel depot-specific adipokine in human adipose tissue: possible role in modulating insulin action. *Am. J. Physiol. Endocrinol. Metab.* **290**, E1253–E1261 (2006).
75. Hotamisligil, G. S. The role of TNF α and TNF receptors in obesity and insulin resistance. *J. Intern. Med.* **245**, 621–625 (1999).
76. Xu, H. *et al.* Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J. Clin. Invest.* **112**, 1821–1830 (2003).
77. Carey, A. L. *et al.* Interleukin-6 increases insulin-stimulated glucose disposal in humans and glucose uptake and fatty acid oxidation *in vitro* via AMP-activated protein kinase. *Diabetes* **55**, 2688–2697 (2006).
78. Rotter, V., Nagaev, I. & Smith, U. Interleukin-6 (IL-6) induces insulin resistance in 3T3-L1 adipocytes and is, like IL-8 and tumor necrosis factor- α , overexpressed in human fat cells from insulin-resistant subjects. *J. Biol. Chem.* **278**, 45777–45784 (2003).
79. Hirosumi, J. *et al.* A central role for JNK in obesity and insulin resistance. *Nature* **420**, 333–336 (2002).
80. Shoelson, S. E., Lee, J. & Yuan, M. Inflammation and the IKK β /I κ B/NF- κ B axis in obesity- and diet-induced insulin resistance. *Int. J. Obes. Relat. Metab. Disord.* **27** (Suppl. 3), S49–S52 (2003).
81. Howard, J. K. & Flier, J. S. Attenuation of leptin and insulin signaling by SOCS proteins. *Trends Endocrinol. Metab.* **17**, 365–371 (2006).
82. Houstis, N., Rosen, E. D. & Lander, E. S. Reactive oxygen species have a causal role in multiple forms of insulin resistance. *Nature* **440**, 944–948 (2006).
83. Steppan, C. M. *et al.* A family of tissue-specific resistin-like molecules. *Proc. Natl Acad. Sci. USA* **98**, 502–506 (2001).
84. Steppan, C. M. *et al.* The hormone resistin links obesity to diabetes. *Nature* **409**, 307–312 (2001).
85. Banerjee, R. R. *et al.* Regulation of fasted blood glucose by resistin. *Science* **303**, 1195–1198 (2004).
86. Steppan, C. M. & Lazar, M. A. The current biology of resistin. *J. Intern. Med.* **255**, 439–447 (2004).
87. Patel, S. D., Rajala, M. W., Rossetti, L., Scherer, P. E. & Shapiro, L. Disulfide-dependent multimeric assembly of resistin family hormones. *Science* **304**, 1154–1158 (2004).
88. Kaser, S. *et al.* Resistin messenger-RNA expression is increased by proinflammatory cytokines *in vitro*. *Biochem. Biophys. Res. Commun.* **309**, 286–290 (2003).
89. Patel, L. *et al.* Resistin is expressed in human macrophages and directly regulated by PPAR γ activators. *Biochem. Biophys. Res. Commun.* **300**, 472–476 (2003).
90. Yang, Q. *et al.* Serum retinol binding protein 4 contributes to insulin resistance in obesity and type 2 diabetes. *Nature* **436**, 356–362 (2005).
91. Graham, T. E. *et al.* Retinol-binding protein 4 and insulin resistance in lean, obese, and diabetic subjects. *N. Engl. J. Med.* **354**, 2552–2563 (2006).
92. Roden, M. *et al.* Mechanism of free fatty acid-induced insulin resistance in humans. *J. Clin. Invest.* **97**, 2859–2865 (1996).
93. Roden, M. *et al.* Effects of free fatty acid elevation on postabsorptive endogenous glucose production and gluconeogenesis in humans. *Diabetes* **49**, 701–707 (2000).
94. Griffin, M. E. *et al.* Free fatty acid-induced insulin resistance is associated with activation of protein kinase C θ and alterations in the insulin signaling cascade. *Diabetes* **48**, 1270–1274 (1999).
95. Schmitz-Peiffer, C. *et al.* Alterations in the expression and cellular localization of protein kinase C isozymes epsilon and theta are associated with insulin resistance in skeletal muscle of the high-fat-fed rat. *Diabetes* **46**, 169–178 (1997).
96. Paolisso, G. *et al.* Does free fatty acid infusion impair insulin action also through an increase in oxidative stress? *J. Clin. Endocrinol. Metab.* **81**, 4244–4248 (1996).
97. Hajdich, E. *et al.* Ceramide impairs the insulin-dependent membrane recruitment of protein kinase B leading to a loss in downstream signalling in L6 skeletal muscle cells. *Diabetologia* **44**, 173–183 (2001).
98. Shi, H. *et al.* TLR4 links innate immunity and fatty acid-induced insulin resistance. *J. Clin. Invest.* **116**, 3015–3025 (2006).
99. Song, M. J., Kim, K. H., Yoon, J. M. & Kim, J. B. Activation of Toll-like receptor 4 is associated with insulin resistance in adipocytes. *Biochem. Biophys. Res. Commun.* **346**, 739–745 (2006).
100. Eldor, R. & Raz, I. Lipotoxicity versus adipotoxicity — the deleterious effects of adipose tissue on beta cells in the pathogenesis of type 2 diabetes. *Diabetes Res. Clin. Pract.* **74**, S3–S8 (2006).
101. Lowell, B. B. & Shulman, G. I. Mitochondrial dysfunction and type 2 diabetes. *Science* **307**, 384–387 (2005).
102. Simha, V. & Garg, A. Lipodystrophy: lessons in lipid and energy metabolism. *Curr. Opin. Lipidol.* **17**, 162–169 (2006).
103. Yamauchi, T. *et al.* The fat-derived hormone adiponectin reverses insulin resistance associated with both lipoatrophy and obesity. *Nature Med.* **7**, 941–946 (2001).
104. Weisberg, S. P. *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *J. Clin. Invest.* **112**, 1796–1808 (2003).
105. Weisberg, S. P. *et al.* CCR2 modulates inflammatory and metabolic effects of high-fat feeding. *J. Clin. Invest.* **116**, 115–124 (2006).
106. Kanda, H. *et al.* MCP-1 contributes to macrophage infiltration into adipose tissue, insulin resistance, and hepatic steatosis in obesity. *J. Clin. Invest.* **116**, 1494–1505 (2006).
107. Kamei, N. *et al.* Overexpression of monocyte chemoattractant protein-1 in adipose tissues causes macrophage recruitment and insulin resistance. *J. Biol. Chem.* **281**, 26602–26614 (2006).
108. Di Gregorio, G. B. *et al.* Expression of CD68 and macrophage chemoattractant protein-1 genes in human adipose and muscle tissues: association with cytokine expression, insulin resistance, and reduction by pioglitazone. *Diabetes* **54**, 2305–2313 (2005).
109. Gnudi, L., Tozzo, E., Shepherd, P. R., Bliss, J. L. & Kahn, B. B. High level overexpression of glucose transporter-4 driven by an adipose-specific promoter is maintained in transgenic mice on a high fat diet, but does not prevent impaired glucose tolerance. *Endocrinology* **136**, 995–1002 (1995).
110. Abel, E. D. *et al.* Adipose-selective targeting of the *GLUT4* gene impairs insulin action in muscle and liver. *Nature* **409**, 729–733 (2001).
111. Fukuchi, K. *et al.* Radionuclide imaging metabolic activity of brown adipose tissue in a patient with pheochromocytoma. *Exp. Clin. Endocrinol. Diabetes* **112**, 601–603 (2004).

Acknowledgements We thank B. Lowell and M. Herman for a critical reading of the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to B.M.S. (bruce_spiegelman@dfci.harvard.edu).

Gut hormones and the regulation of energy homeostasis

Kevin G. Murphy¹ & Stephen R. Bloom¹

Food intake, energy expenditure and body adiposity are homeostatically regulated. Central and peripheral signals communicate information about the current state of energy balance to key brain regions, including the hypothalamus and brainstem. Hunger and satiety represent coordinated responses to these signals, which include neural and hormonal messages from the gut. In recent years our understanding of how neural and hormonal brain–gut signalling regulates energy homeostasis has advanced considerably. Gut hormones have various physiological functions that include specifically targeting the brain to regulate appetite. New research suggests that gut hormones can be used to specifically regulate energy homeostasis in humans, and offer a target for anti-obesity drugs.

Energy balance is a homeostatic system. Although malfunctions of this system can cause obesity¹, the relatively recent increase in the incidence of obesity is not thought to be the result of specific defects, but of a regulatory system unable to cope with the current context of cheap, high-energy foodstuffs, mechanized transport and non-manual labour. Commandeering elements of this regulatory system might provide the best opportunity for us to combat obesity.

The main brain regions responsible for the regulation of energy homeostasis are the hypothalamus and the brainstem. A number of current obesity therapies target central neurotransmitters in order to reduce body weight, as did several past treatments. Sibutramine reduces appetite by potentiating noradrenaline and serotonin signalling, and rimonabant achieves the same goal by antagonizing central type-1 cannabinoid receptors. Sibutramine and rimonabant both produced a modest reduction in body weight in clinical trials². However, the neurotransmitter systems affected by such drugs do not exclusively regulate appetite, so there is a high likelihood of side effects. For example, sibutramine can cause hypertension³, and there are concerns that rimonabant might cause psychological and reproductive problems^{4,5}.

The hypothalamus and brainstem receive neural and hormonal signals from the periphery that encode information about acute nutritional state and adiposity. The adipose hormone leptin signals to the brain the size of adipose stores but does not show great promise as an anti-obesity agent. Most obese humans already have high circulating levels of leptin and are resistant to its metabolic effects⁶. So far, no means of preventing this resistance has been validated.

However, shorter-term signals can influence food intake. Postprandial satiety and the hunger felt before a meal are not thought to be primarily mediated by leptin. Neural and endocrine signalling from the gut are believed to have important roles in this short-term regulation of appetite. Mechanoreceptors and chemoreceptors in the gastrointestinal tract signal through the vagal nerve to the brainstem. These neural signals are centrally integrated with those transmitted by a number of hormones released from the gastrointestinal tract and its associated structures. These gut hormones then stimulate ascending vagal pathways from the gut to the brainstem or act directly on neurons in the brain.

This review examines the role of specific gut hormones in the regulation of energy homeostasis. We conclude that gut hormones have physi-

ological and pathophysiological roles in appetite regulation, and might represent useful targets for future obesity therapies.

Gut hormones

The gastrointestinal tract is the body's largest endocrine organ and releases more than 20 different regulatory peptide hormones that influence a number of physiological processes and act on tissues including exocrine glands, smooth muscle and the peripheral nervous system. Most of these hormones are sensitive to gut nutrient content, and short-term feelings of hunger and satiety are believed to be mediated, in part, by coordinated changes in circulating gut hormone levels⁷.

Ghrelin

Ghrelin is a peptide hormone released into circulation from the stomach that was first discovered as an endogenous ligand for the growth-hormone-secretaagogue receptor (GHS-R). Ghrelin is composed of 28 amino acids and is uniquely modified by the addition of an octanoyl group to the serine residue at position three. This acylation is necessary for ghrelin to bind to the GHS-R and to cross the blood–brain barrier⁸. Great excitement heralded reports that ghrelin increased appetite. Although a number of circulating factors, including the adipose hormone leptin and several gut hormones, have been reported to reduce food intake, ghrelin is the only known factor to increase appetite through the circulation. The pattern of ghrelin release suggests that it governs feelings of hunger. Circulating ghrelin levels are increased by fasting, and fall after a meal. Central or peripheral administration of acylated ghrelin to rats acutely stimulates food intake and growth hormone release, and chronic administration causes weight gain. Intravenous infusion or subcutaneous injection of ghrelin to humans increases both feelings of hunger and food intake. Ghrelin is often referred to as the 'hunger hormone'⁸.

It has been reported that peripheral ghrelin administration reduces fat use⁹ and that chronic central ghrelin infusion increases the expression of enzymes that promote fat storage in adipose tissue¹⁰. The metabolic effects of ghrelin might not, therefore, be entirely dependent on increased food intake.

Initial reports suggested that *Ghs-r*-knockout and ghrelin-knockout mice lacked important alterations in feeding behaviours and adiposity. However, it has since been discovered that models of disrupted ghrelin

¹Department of Metabolic Medicine, Imperial College Faculty of Medicine, Hammersmith Campus, Du Cane Road, London W12 0NN, UK.

signalling can show complex energy homeostasis-related phenotypes. *Ghs-r*-knockout mice and male ghrelin-deficient mice show resistance to diet-induced obesity when supplied with a high-fat diet early in life. Ghrelin-deficient mice also show altered metabolic fuel preference, favouring fat as a source of energy to a greater extent than wild-type controls when fed a high fat diet¹¹. Other studies have found that diabetic ghrelin-knockout mice show attenuated hyperphagia¹², and that loss of ghrelin attenuates diabetes in the *ob/ob* obese mouse¹³.

The idea of blocking ghrelin signalling with GHS-R antagonists is of interest as a means of preventing obesity. A GHS-R antagonist has been reported to reduce food intake in fasted mice and an RNA Spiegelmer (an L-oligonucleotide designed to bind specifically to a particular molecule) to inhibit ghrelin action *in vitro* and *in vivo*¹¹. Recently, it has been demonstrated that vaccinating rats against ghrelin can reduce weight gain¹⁴. However, it may be unwise to use a similarly irreversible technology in humans, and it remains to be shown whether blocking ghrelin signalling is a viable long-term obesity therapy for humans.

Ghrelin agonists might be useful in the treatment of specific patient groups with anorexia. Acute ghrelin administration can stimulate appetite in cancer patients with appetite loss¹⁵ and in dialysis patients with malnutrition¹⁶, and has also been shown to increase gastric emptying in diabetic patients with gastroparesis¹⁷. Thus anorectic patients might benefit from the orexigenic effects of chronic ghrelin administration, but this remains to be demonstrated. The question of possible side effects would also need to be addressed before ghrelin could be used as a therapy¹⁸. Interestingly, it has recently been reported that peripheral ghrelin administration can also modify hypothalamic plasticity¹⁹ and hippocampal neuron formation²⁰.

The gene that encodes ghrelin has also been found to encode another peptide known as obestatin. Obestatin was originally reported to reduce food intake when administered peripherally or intracerebroventricularly, and to reduce body weight gain when administered peripherally²¹. These effects were proposed to be mediated by the orphan G-protein-coupled receptor, GPR39. Much speculation followed as to why the same gene would produce an orexigenic and an anorectic signal. However, subsequent reports have not supported the initial findings and suggest that obestatin might not signal through GPR39 or have a role in the regulation of food intake^{22–24}.

Peptide YY

Peptide YY (PYY) is a gut hormone that is related to neuropeptide Y (NPY). Both peptides have the PP fold structural motif and exert their effects through the Y family of receptors. Full-length PYY binds with similar affinity to all of the Y receptors. However, most circulating PYY-immunoreactivity is in the amino-terminally truncated form, PYY_{3–36}, which preferentially binds to the Y2 receptor (Y2R).

PYY is found in L cells throughout the length of the gut, although it is present at higher concentrations in more distal sections. PYY is released into the circulation after a meal and is reduced by fasting. Acute peripheral administration of PYY_{3–36} reduces food intake in rodents and humans²⁵. These findings were initially contentious, with a number of independent laboratories being unable to repeat aspects of the study²⁶. Stress can reduce baseline food intake, making it difficult for anorectic agents to further suppress appetite. Peripheral PYY_{3–36} administration does not reduce food intake in rodents not acclimatized to experimental procedures, or in rats presented with a new environment^{27,28}. The susceptibility of the anorectic effects of PYY_{3–36} to these mild stressors might explain the difficulties others initially had in replicating the original PYY_{3–36} results. Subsequent studies have confirmed that PYY_{3–36} acutely inhibits feeding in rodents and primates^{29–34}. The anorectic effects of PYY_{3–36} are thought to be mediated by Y2R, because they are attenuated by Y2R antagonists³⁵ and abolished in Y2R-knockout mice²⁵. *Pyy*-knockout mice have disrupted energy homeostasis, suggesting that the PYY system does have a physiological role in its regulation^{36,37}.

Obese humans show normal sensitivity to the anorectic effects of PYY_{3–36}, and circulating levels of PYY are not raised in the obese, in contrast to leptin levels³⁸. Some studies have reported that obese indi-

viduals have lower fasting and postprandial levels of circulating PYY^{38–40} but others have found no significant differences between PYY levels in the lean and the obese^{41,42}, suggesting that a reduction in PYY release is unlikely to be involved in the aetiology of obesity.

High doses of PYY_{3–36} have been reported to cause conditioned taste aversion in animals^{43,44} and nausea in humans⁴⁵. This effect is worsened by rapid administration. Steady low-dose intravenous infusion of PYY_{3–36} can reduce food intake without aversive effects in rats⁴⁴ or nausea in humans^{25,38,40}.

The efficacy of chronically administered PYY_{3–36} in reducing food intake is vital to its use as a possible obesity drug. Different chronic administration protocols have drawn different conclusions^{25,26}. Intravenous PYY_{3–36} administration has more consistently supported its anorectic effects, suggesting that administrative route is an issue^{31,46}. It has recently been shown that intermittent intravenous administration of PYY_{3–36} can cause long-term reductions in food intake, body weight and adiposity in rats. However, dosage pattern seems to be crucial in producing sustained reductions in food intake and body weight⁴⁷.

Cholecystokinin

The role of cholecystokinin in the exocrine pancreas and gallbladder was long established when, in 1973, it became the first gut hormone to be shown to influence food intake. Subsequent studies confirmed these findings in rodents and humans. Cholecystokinin is released postprandially from the small intestine, and seems to reduce food intake through cholecystokinin 1 (CCK1) receptors on the vagal nerve. CCK1 receptor antagonists have been reported to increase food intake in rodents and humans, and the Otsuka Long-Evans Tokushima Fatty rat, which lacks the CCK1 receptor, is hyperphagic and obese, suggesting that cholecystokinin has a physiological role in the regulation of food intake⁴⁸. However, continuous infusion of cholecystokinin fails to influence food intake after 24 hours, and although intermittent administration does reduce food intake acutely, this effect is compensated for by increased food intake between injections^{49–51}. It may be that a stringent administration regimen is required to reduce body weight. It has been reported that cholecystokinin cannot activate the appropriate vagal circuits at endogenous circulating concentrations, suggesting that the actions of cholecystokinin on food intake might be paracrine or neurocrine rather than endocrine⁵². High doses of cholecystokinin cause nausea, but lower doses, which do not, can still reduce food intake. Hunger, satiety and nausea have been suggested to represent points on the same physiological spectrum⁵³. Nausea is associated with high-dose administration of all of the anorectic gut hormones and analogues, including PYY_{3–36} (ref. 45), glucagon-like peptide-1, exendin-4 (ref. 54) and oxyntomodulin⁵⁵.

Pancreatic polypeptide

Pancreatic polypeptide is a member of the PP fold peptide family that is synthesized and released from the endocrine pancreas. Pancreatic polypeptide shows preferential binding to the Y4 and Y5 receptors. Similarly to PYY_{3–36}, it is released after a meal and reduces appetite. Acute peripheral administration to mice and humans reduces food intake^{56,57}, and chronic administration reduces body weight in *ob/ob* obese mice⁵⁸. No nausea or gastrointestinal distress was reported in the intravenous study investigating feeding effects in humans. The anorectic effects of pancreatic polypeptide might occur partly as a result of delayed gastric emptying^{57,59}. The food intake and fat mass of transgenic mice overexpressing pancreatic polypeptide are reduced, but such mice also show reduced gastric emptying⁵⁶. However, at present the physiological role of pancreatic polypeptide in energy homeostasis is unknown.

Amylin

Amylin, which is also known as islet amyloid polypeptide, is a 37-residue member of the calcitonin peptide family that is released together with insulin from pancreatic β -cells in response to food ingestion. Although its main function is thought to be in glucose homeostasis,

given peripherally at supraphysiological levels amylin can reduce food intake. Administration of the amylin agonist pramlintide reduces body weight in type 1 and 2 diabetics by between 0.5 and 1.4 kg for up to 1 year^{60,61}.

Glucose-dependent insulintropic polypeptide

Glucose-dependent insulintropic polypeptide (GIP) is a 42-residue peptide released from K cells in the duodenum after food ingestion. GIP has not been reported to have an acute influence on food intake. However, GIP-receptor-knockout mice are resistant to obesity when fed a high-fat diet. The reason for this resistance is unclear and there has been speculation that it might reflect a direct effect on adipocytes rather than on central appetite-regulating circuits^{62,63}.

Glucagon-like peptide-1

The same gut endocrine cell type that synthesizes PYY also synthesizes a large precursor protein known as preproglucagon. This is processed further to produce a number of biologically active peptides, including glucagon, glucagon-like peptide-1, glucagon-like peptide-2 (GLP-2) and oxyntomodulin.

Glucagon-like peptide-1 exists in several forms, but the most common circulating form is glucagon-like peptide-1_{7–36amide} (GLP-1). GLP-1 is released into the circulation after a meal and is a potent incretin — central or peripheral administration strongly stimulates insulin release. Intracerebroventricular administration of GLP-1 potently reduces food intake in rodents, and peripheral administration of GLP-1 inhibits appetite in animals and humans⁵⁴.

The saliva of the Gila monster lizard, *Heloderma suspectum*, contains a peptide known as exendin-4, which is a potent GLP-1 receptor agonist. A truncated form of this peptide, exendin 9–39, acts as a competitive antagonist at the same receptor. Acute intracerebroventricular administration of exendin 9–39 increases food intake, and chronic administration increases body weight. It seems, therefore, that endogenous peripheral GLP-1 might form part of the physiological mechanism to reduce appetite and food intake after a meal. However, the food intake and body weight of GLP-1-receptor-knockout mice are normal⁵⁴.

Clinical trials have shown that exendin-4 (also known as exenatide, and marketed as Byetta) is useful in the regulation of glucose homeostasis in people with type 2 diabetes mellitus. Interestingly, in phase III 30-week clinical trials, exenatide significantly reduced body weight in treated diabetics^{64–66}. Although nausea is a relatively common side effect of the treatment, as with PYY_{3–36}, it does not seem to be intrinsically linked to the effects on appetite⁵⁴. These results also suggest that even if peripheral GLP-1 does not physiologically regulate appetite, the GLP-1 system might be used to reduce body weight using a peripherally administered drug.

Glucagon-like peptide-2

GLP-2 is found in the brain and inhibits food intake when administered centrally. Peripheral circulating GLP-2, however, is primarily involved in stimulating gastrointestinal motility, absorption and growth, and does not influence appetite physiologically⁶⁷.

Oxyntomodulin

Oxyntomodulin is also a product of the preproglucagon precursor molecule and, like GLP-1, is released after a meal. Also like GLP-1, it reduces food intake when centrally administered to rats and peripherally administered to rodents and humans^{68,69}. This is not surprising — oxyntomodulin seems to signal through the GLP-1 receptor: its anorectic effects are blocked by exendin 9–39 (ref. 68) and abolished in GLP-1-receptor-knockout mice⁷⁰. It also causes a similar pattern of neuronal activation to GLP-1 after peripheral administration⁷⁰.

However, there is evidence that oxyntomodulin does not merely mirror the activities of GLP-1. Oxyntomodulin has a roughly 50-fold lower affinity for the GLP-1 receptor than GLP-1, but seems to reduce food intake with similar potency⁷¹. Although the administration of exendin 9–39 directly into the hypothalamic arcuate nucleus (ARC) has been reported to block the anorectic effects of oxyntomodulin, it does not block those of GLP-1 (ref. 68). Oxyntomodulin and GLP-1 might therefore have different roles in energy homeostasis, perhaps mediated by their different pharmacological properties or by tissue-specific signalling factors.

Chronic central or peripheral oxyntomodulin administration reduces weight gain in rats^{68,72}. In addition, chronic oxyntomodulin treatment causes rats to lose more weight than pair-fed controls, suggesting that oxyntomodulin increases energy expenditure⁷². Chronic administration of oxyntomodulin can also cause weight loss in humans. In a 4-week study in which oxyntomodulin or saline were self-administered by overweight and obese volunteers, the oxyntomodulin-treated group demonstrated an average weight loss of 0.45 kg per week more than the saline-treated group⁵⁵. As in rats, oxyntomodulin might increase energy expenditure in humans. Self-administration of oxyntomodulin by overweight and obese volunteers for 4 days significantly increased activity-related energy expenditure, as assessed by continuous electronic movement monitoring⁷³. Further work is required to investigate whether oxyntomodulin is effective over long administration periods.

Gut hormones signal to central appetite circuits

Gut hormones can activate circuits in the hypothalamus and brainstem, the main central nervous system centres responsible for the regulation of energy homeostasis. For many gut hormones, the precise mechanisms of central action are unknown or contentious. A number of gut hormones also act as neurotransmitters in the brain, where they do not necessarily serve the same functions as in the periphery, making it difficult to tease out their endocrine effects.

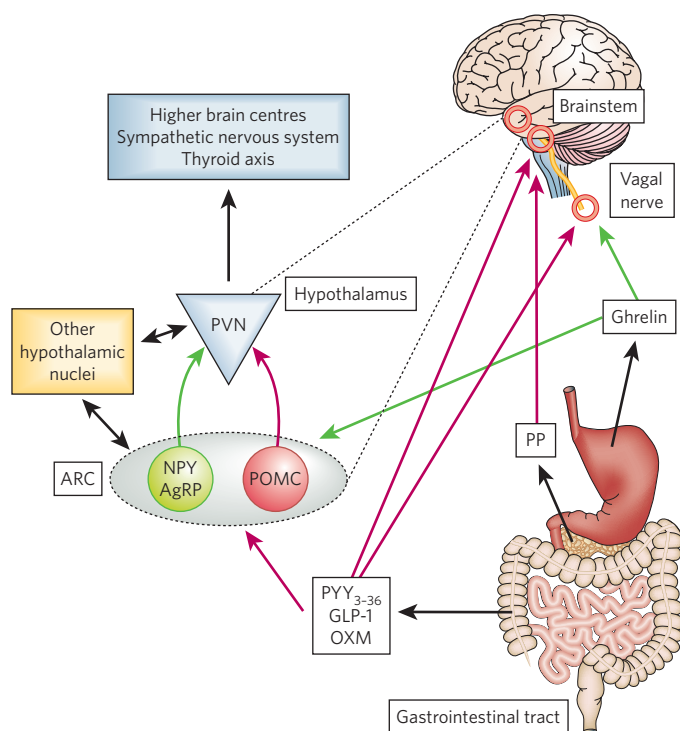


Figure 1 | The pathways by which gut hormones regulate energy homeostasis. PYY_{3–36}, GLP-1 and oxyntomodulin (OXM) are released from gut L cells after a meal. They can directly stimulate anorectic pathways in the hypothalamus and brainstem, and may also act through the vagus nerve. Pancreatic polypeptide (PP) is released after a meal and is thought to reduce appetite by directly signalling to neurons in the brainstem. Ghrelin is released from the stomach with fasting and might signal directly to the hypothalamus or through the vagus nerve to stimulate food intake. The ARC is important in integrating gut hormone energy homeostasis signals. NPY/AgRP neurons and POMC neurons signal to the PVN and other hypothalamic nuclei to increase or decrease appetite, respectively. Green arrows indicate orexigenic signals and red arrows indicate anorectic signals.

The ARC is believed to be important in integrating peripheral circulating energy homeostasis signals. ARC neurons expressing NPY and agouti-related peptide (AgRP) signal to increase appetite, and pro-opiomelanocortin (POMC)-expressing melanocortin neurons signal to reduce appetite. The hypothalamic paraventricular nucleus (PVN) is thought to be a critical target of these ARC neurons. The PVN signals to higher brain centres and the sympathetic nervous system, and regulates the thyroid axis. ARC neurons also project to other hypothalamic nuclei that signal, in turn, to the ARC and the PVN to modulate their activity⁶ (Fig. 1).

Ghrelin activates NPY neurons in the ARC, and blocking NPY or agouti-related protein (AgRP) signalling abolishes ghrelin's orexigenic actions⁸. Early evidence suggested that PYY₃₋₃₆ reduced food intake through ARC melanocortin neurons^{25,27,30}. However, PYY₃₋₃₆ still reduces food intake in mice with disrupted melanocortin signalling, and has been reported to inhibit melanocortin neurons in further experiments, suggesting that PYY₃₋₃₆ reduces food intake by another, as yet unknown, mechanism^{27,30,74,75}.

There are extensive reciprocal connections between the hypothalamus and the brainstem, and energy intake is coordinated on the basis of information received by both regions^{6,54,76}. In the brainstem, the nucleus of the solitary tract (NTS), area postrema and dorsal motor nucleus of the vagus have all been implicated in the regulation of energy homeostasis. Peripheral GLP-1 administration activates neurons in the brainstem and in the hypothalamus^{54,70,77}. Peripheral PYY₃₋₃₆ administration has been reported to increase c-Fos immunoreactivity in the area postrema and the NTS. Y2R agonists can inhibit excitatory postsynaptic currents in NTS neurons *in vitro*⁷⁸, and in dorsal motor nucleus neurons *in vivo* and *in vitro*⁷⁹. In addition, transecting brainstem–hypothalamic pathways in rodents blocks PYY₃₋₃₆-induced anorexia⁸⁰. The brainstem is therefore likely to have some role in PYY₃₋₃₆ signalling. Pancreatic polypeptide has been suggested to act by directly activating Y4-receptor-expressing neurons in the area postrema⁵⁶.

Gut hormones can signal to the brainstem through the vagal nerve. There is strong evidence that this is how cholecystokinin mediates its effects on food intake⁵². The physiological relevance of vagal signalling to other gut hormone systems is less certain. Vagotomy has been reported to block or attenuate the orexigenic effects of ghrelin^{81,82} and the anorectic effects of PYY₃₋₃₆ (refs 77, 80), GLP-1 (ref. 77) and pancreatic polypeptide⁵⁸. However, blocking vagal signalling alters the regulation of energy homeostasis and gastrointestinal function considerably, so it is difficult to prove that the effects of vagotomy are specific.

Hypothalamic, brainstem and vagal signalling all have a role in appetite control, but further research is required to determine how each of these signals is weighted and integrated.

The role of gut hormones in energy homeostasis

Gut hormones have a number of functions, including the regulation of blood glucose levels, gastrointestinal motility and growth, exocrine secretion and adipocyte function (Fig. 2). These functions are often integrated with their actions in the central regulation of appetite circuits, and the gut hormones themselves interact to stimulate or suppress the release of other hormones. For example, cholecystokinin stimulates PYY release⁸³, whereas oxyntomodulin, PYY₃₋₃₆ and insulin suppress ghrelin levels^{25,68,69,84}. Interestingly, there is evidence that insulin and glucagon can suppress ghrelin through central mechanisms^{85,86}. Although the system is complex, by examining the wide-ranging effects of gut hormones on appetite it is possible to discern simple themes.

There seem to be three major roles for gut hormones in appetite regulation. First, the release of gut hormones can modulate normal hunger and satiety. Circulating ghrelin levels increase before a meal and correspond to hunger pangs. A number of anorectic gut hormones are released postprandially as satiety signals. However, these increases in circulating concentrations are often small, and it seems likely that satiety might represent the cumulative effects of a number of submaximal gut hormone responses. Gut hormones might thus have additive effects on appetite^{87,88}.

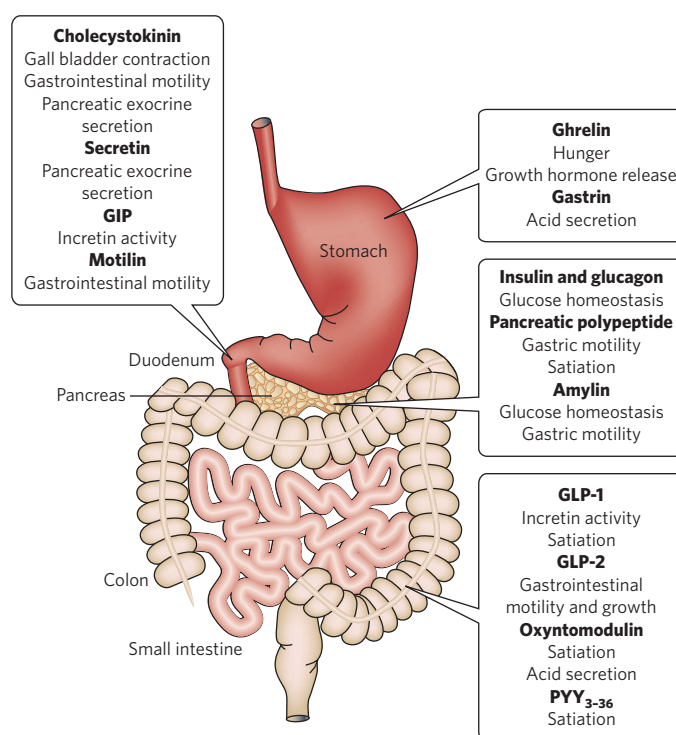


Figure 2 | A schematic diagram of the gastrointestinal tract illustrating where particular gut hormones are concentrated and their major putative functions. The gastrointestinal tract releases a number of hormones, including ghrelin and gastrin from the stomach, insulin, glucagon, pancreatic polypeptide and amylin from the pancreas, cholecystokinin, secretin, GIP and motilin from the small intestine, and GLP-1, GLP-2, oxyntomodulin and PYY₃₋₃₆ from the large intestine. These hormones signal to the periphery and to the central nervous system to regulate a number of biological processes.

Second, gut hormones may reduce food intake in patients with specific gut diseases, perhaps as an adaptation to reduce further stress on the gut. A number of anorectic gut hormones are elevated in gut disease⁸⁹⁻⁹¹. This might be a specific function of gut hormones in the distal gastrointestinal tract. Enteroendocrine cells might, for example, release gut hormones at high level in response to undigested foodstuffs, the presence of which would suggest that the function of the upper gastrointestinal tract is compromised. Consistent with this hypothesis, gut bypass surgery is associated with elevated anorectic gut hormone concentrations, which are believed to be at least partly responsible for the reduction in appetite in gut bypass patients^{39,92-95}.

Third, very high levels of gut hormones may be released to generate conditioned taste aversion and nausea in response to the ingestion of harmful substances. The seeming profligacy of the L cell, which releases three distinct anorectic hormones (PYY₃₋₃₆, GLP-1 and oxyntomodulin), makes more sense if we consider the complex information these hormones might encode about short-term energy balance and the state of the gastrointestinal tract.

The future of gut hormones in appetite control

Current drugs are insufficiently efficacious to cope with the obesity epidemic sweeping the developed world. At present, the most effective treatment for obesity is bariatric surgery. However, its cost and associated mortality make it impractical to treat rising numbers of obese patients, and it is increasingly reserved for only the morbidly obese. Gut hormones are molecules designed by evolution to be 'administered' peripherally to specifically target appetite circuits in the central nervous system. Although the physiological role of a particular gut hormone in energy homeostasis can be difficult to prove conclusively, administration of exogenous gut hormones can certainly influence food intake in

humans. Hijacking such systems to tackle obesity might prove effective, even if gut hormones do not significantly regulate appetite on a day-to-day basis.

The physiological and pathophysiological roles of gut hormones in energy balance therefore remain to be defined. However, as the obesity epidemic rumbles on, so do the efforts to sate hunger. Gut hormones may yet prove that the way to a man's brain is through his stomach. ■

- Farooqi, I. S. & O'Rahilly, S. Monogenic obesity in humans. *Annu. Rev. Med.* **56**, 443–458 (2005).
- Halford, J. C. Pharmacotherapy for obesity. *Appetite* **46**, 6–10 (2006).
- James, W. P. *et al.* Effect of sibutramine on weight maintenance after weight loss: a randomised trial. STORM Study Group. Sibutramine Trial of Obesity Reduction and Maintenance. *Lancet* **356**, 2119–2125 (2000).
- Gelfand, E. V. & Cannon, C. P. Rimonabant: a cannabinoid receptor type 1 blocker for management of multiple cardiometabolic risk factors. *J. Am. Coll. Cardiol.* **47**, 1919–1926 (2006).
- Wang, H., Dey, S. K. & Maccarrone, M. Jekyll and Hyde: two faces of cannabinoid signaling in male and female fertility. *Endocr. Rev.* **27**, 427–448 (2006).
- Schwartz, M. W., Woods, S. C., Porte, D., Seeley, R. J. & Baskin, D. G. Central nervous system control of food intake. *Nature* **404**, 661–671 (2000).
- Badman, M. K. & Flier, J. S. The gut and energy balance: visceral allies in the obesity wars. *Science* **307**, 1909–1914 (2005).
- Kojima, M. & Kangawa, K. Ghrelin: structure and function. *Physiol. Rev.* **85**, 495–522 (2005).
- Tschöp, M., Smiley, D. L. & Heiman, M. L. Ghrelin induces adiposity in rodents. *Nature* **407**, 908–913 (2000).
- Theander-Carrillo, C. *et al.* Ghrelin action in the brain controls adipocyte metabolism. *J. Clin. Invest.* **116**, 1983–1993 (2006).
- Cummings, D. E. Ghrelin and the short- and long-term regulation of appetite and body weight. *Physiol. Behav.* **89**, 71–84 (2006).
- Dong, J. *et al.* Role of endogenous ghrelin in the hyperphagia of mice with streptozotocin-induced diabetes. *Endocrinology* **147**, 2634–2642 (2006).
- Sun, Y., Asnicar, M., Saha, P. K., Chan, L. & Smith, R. G. Ablation of ghrelin improves the diabetic but not obese phenotype of *ob/ob* mice. *Cell Metab.* **3**, 379–386 (2006).
- Zorrilla, E. P. *et al.* Vaccination against weight gain. *Proc. Natl Acad. Sci. USA* (2006).
- Neary, N. M. *et al.* Ghrelin increases energy intake in cancer patients with impaired appetite: acute, randomized, placebo-controlled trial. *J. Clin. Endocrinol. Metab.* **89**, 2832–2836 (2004).
- Wynne, K. *et al.* Subcutaneous ghrelin enhances acute food intake in malnourished patients who receive maintenance peritoneal dialysis: a randomized, placebo-controlled trial. *J. Am. Soc. Nephrol.* **16**, 2111–2118 (2005).
- Murray, C. D. *et al.* Ghrelin enhances gastric emptying in diabetic gastroparesis: a double blind, placebo controlled, crossover study. *Gut* **54**, 1693–1698 (2005).
- Nogueiras, R., Perez-Tilve, D., Wortley, K. E. & Tschöp, M. Growth hormone secretagogue (ghrelin-) receptors — a complex drug target for the regulation of body weight. *CNS Neurol. Disord. Drug Targets* **5**, 335–343 (2006).
- Pinto, S. *et al.* Rapid rewiring of arcuate nucleus feeding circuits by leptin. *Science* **304**, 110–115 (2004).
- Diano, S. *et al.* Ghrelin controls hippocampal spine synapse density and memory performance. *Nature Neurosci.* **9**, 381–388 (2006).
- Zhang, J. V. *et al.* Obestatin, a peptide encoded by the ghrelin gene, opposes ghrelin's effects on food intake. *Science* **310**, 996–999 (2005).
- Gourcerol, G. *et al.* Lack of interaction between peripheral injection of CCK and obestatin in the regulation of gastric satiety signaling in rodents. *Peptides* **27**, 2811–2819 (2006).
- Holst, B. *et al.* GPR39 signaling is stimulated by zinc ions but not by obestatin. *Endocrinology* advance online publication, 7 September 2006 (doi:10.1210/en.2006-0933).
- Nogueiras, R. *et al.* Effects of obestatin on energy balance and growth hormone secretion in rodents. *Endocrinology* advance online publication, 28 September 2006 (doi:10.1210/en.2006-0915).
- Batterham, R. L. *et al.* Gut hormone PYY_{3–36} physiologically inhibits food intake. *Nature* **418**, 650–654 (2002).
- Tschöp, M. *et al.* Physiology: does gut hormone PYY_{3–36} decrease food intake in rodents? *Nature* **430**, 1 (2004).
- Halatchev, I. G., Ellacott, K. L., Fan, W. & Cone, R. D. Peptide YY_{3–36} inhibits food intake in mice through a melanocortin-4 receptor-independent mechanism. *Endocrinology* **145**, 2585–2590 (2004).
- Abbott, C. R. *et al.* The importance of acclimatisation and habituation to experimental conditions when investigating the anorectic effects of gastrointestinal hormones in the rat. *Int. J. Obes. (Lond.)* **30**, 288–292 (2006).
- Adams, S. H., Won, W. B., Schonhoff, S. E., Leiter, A. B. & Paterniti, J. R. Effects of peptide YY_{3–36} on short-term food intake in mice are not affected by prevailing plasma ghrelin levels. *Endocrinology* **145**, 4967–4975 (2004).
- Challis, B. G. *et al.* Mice lacking pro-opiomelanocortin are sensitive to high-fat feeding but respond normally to the acute anorectic effects of peptide-YY_{3–36}. *Proc. Natl Acad. Sci. USA* **101**, 4695–4700 (2004).
- Chelikani, P. K., Haver, A. C. & Reidelberger, R. D. Intravenous infusion of peptide YY_{3–36} potently inhibits food intake in rats. *Endocrinology* **146**, 879–888 (2005).
- Cox, J. E. & Randich, A. Enhancement of feeding suppression by PYY_{3–36} in rats with area postrema ablations. *Peptides* **25**, 985–989 (2004).
- Pittner, R. A. *et al.* Effects of PYY_{3–36} in rodent models of diabetes and obesity. *Int. J. Obes. Relat. Metab. Disord.* **28**, 963–971 (2004).
- Moran, T. H. *et al.* Peptide YY_{3–36} inhibits gastric emptying and produces acute reductions in food intake in rhesus monkeys. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **288**, R384–R388 (2005).
- Abbott, C. R. *et al.* Blockade of the neuropeptide Y Y2 receptor with the specific antagonist BIIIE0246 attenuates the effect of endogenous and exogenous peptide YY_{3–36} on food intake. *Brain Res.* **1043**, 139–144 (2005).
- Boey, D. *et al.* Peptide YY ablation in mice leads to the development of hyperinsulinaemia and obesity. *Diabetologia* **49**, 1360–1370 (2006).
- Batterham, R. L. *et al.* Critical role for peptide YY in protein-mediated satiation and body-weight regulation. *Cell Metab.* **4**, 223–233 (2006).
- Batterham, R. L. *et al.* Inhibition of food intake in obese subjects by peptide YY_{3–36}. *N. Engl. J. Med.* **349**, 941–948 (2003).
- Korner, J. *et al.* Effects of Roux-en-Y gastric bypass surgery on fasting and postprandial concentrations of plasma ghrelin, peptide YY, and insulin. *J. Clin. Endocrinol. Metab.* **90**, 359–365 (2005).
- le Roux, C. W. *et al.* Attenuated peptide YY release in obese subjects is associated with reduced satiety. *Endocrinology* **147**, 3–8 (2006).
- Kim, B. J. *et al.* Peptide YY is secreted after oral glucose administration in a gender-specific manner. *J. Clin. Endocrinol. Metab.* **90**, 6665–6671 (2005).
- Stock, S. *et al.* Ghrelin, peptide YY, glucose-dependent insulinotropic polypeptide, and hunger responses to a mixed meal in anorexic, obese, and control female adolescents. *J. Clin. Endocrinol. Metab.* **90**, 2161–2168 (2005).
- Halatchev, I. G. & Cone, R. D. Peripheral administration of PYY_{3–36} produces conditioned taste aversion in mice. *Cell Metab.* **1**, 159–168 (2005).
- Chelikani, P. K., Haver, A. C. & Reidelberger, R. D. Dose-dependent effects of peptide YY_{3–36} on conditioned taste aversion in rats. *Peptides* advance online publication 6 September 2006 (doi:10.1016/j.peptides.2006.08.001).
- Degen, L. *et al.* Effect of peptide YY_{3–36} on food intake in humans. *Gastroenterology* **129**, 1430–1436 (2005).
- Scott, V., Kimura, N., Stark, J. A. & Luckman, S. M. Intravenous peptide YY_{3–36} and Y2 receptor antagonism in the rat: effects on feeding behaviour. *J. Neuroendocrinol.* **17**, 452–457 (2005).
- Chelikani, P. K., Haver, A. C., Reeve, J. R., Keire, D. A. & Reidelberger, R. D. Daily, intermittent intravenous infusion of peptide YY_{3–36} reduces daily food intake and adiposity in rats. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **290**, R298–R305 (2006).
- Bi, S. & Moran, T. H. Actions of CCK in the controls of food intake and body weight: lessons from the CCK-A receptor deficient OLETF rat. *Neuropeptides* **36**, 171–181 (2002).
- Crawley, J. N. & Beinfeld, M. C. Rapid development of tolerance to the behavioural actions of cholecystokinin. *Nature* **302**, 703–706 (1983).
- West, D. B., Fey, D. & Woods, S. C. Cholecystokinin persistently suppresses meal size but not food intake in free-feeding rats. *Am. J. Physiol.* **246**, R776–R787 (1984).
- West, D. B. *et al.* Infusion of cholecystokinin between meals into free-feeding rats fails to prolong the intermeal interval. *Physiol. Behav.* **39**, 111–115 (1987).
- Moran, T. H. Cholecystokinin and satiety: current perspectives. *Nutrition* **16**, 858–865 (2000).
- Greenough, A., Cole, G., Lewis, J., Lockton, A. & Blundell, J. Untangling the effects of hunger, anxiety, and nausea on energy intake during intravenous cholecystokinin octapeptide (CCK-8) infusion. *Physiol. Behav.* **65**, 303–310 (1998).
- Drucker, D. J. The biology of incretin hormones. *Cell Metab.* **3**, 153–165 (2006).
- Wynne, K. *et al.* Subcutaneous oxyntomodulin reduces body weight in overweight and obese subjects: a double-blind, randomized, controlled trial. *Diabetes* **54**, 2390–2395 (2005).
- Katsuura, G., Asakawa, A. & Inui, A. Roles of pancreatic polypeptide in regulation of food intake. *Peptides* **23**, 323–329 (2002).
- Batterham, R. L. *et al.* Pancreatic polypeptide reduces appetite and food intake in humans. *J. Clin. Endocrinol. Metab.* **88**, 3989–3992 (2003).
- Asakawa, A. *et al.* Characterization of the effects of pancreatic polypeptide in the regulation of energy balance. *Gastroenterology* **124**, 1325–1336 (2003).
- Schmidt, P. T. *et al.* A role for pancreatic polypeptide in the regulation of gastric emptying and short-term metabolic control. *J. Clin. Endocrinol. Metab.* **90**, 5241–5246 (2005).
- Whitehouse, F. *et al.* A randomized study and open-label extension evaluating the long-term efficacy of pramlintide as an adjunct to insulin therapy in type 1 diabetes. *Diabetes Care* **25**, 724–730 (2002).
- Ratner, R. *et al.* Adjunctive therapy with pramlintide lowers HbA1c without concomitant weight gain and increased risk of severe hypoglycemia in patients with type 1 diabetes approaching glycemic targets. *Exp. Clin. Endocrinol. Diabetes* **113**, 199–204 (2005).
- Miyawaki, K. *et al.* Inhibition of gastric inhibitory polypeptide signaling prevents obesity. *Nature Med.* **8**, 738–742 (2002).
- Miyawaki, K. *et al.* Glucose intolerance caused by a defect in the entero-insular axis: a study in gastric inhibitory polypeptide receptor knockout mice. *Proc. Natl Acad. Sci. USA* **96**, 14843–14847 (1999).
- Buse, J. B. *et al.* Effects of exenatide (exendin-4) on glycemic control over 30 weeks in sulfonylurea-treated patients with type 2 diabetes. *Diabetes Care* **27**, 2628–2635 (2004).
- DeFronzo, R. A. *et al.* Effects of exenatide (exendin-4) on glycemic control and weight over 30 weeks in metformin-treated patients with type 2 diabetes. *Diabetes Care* **28**, 1092–1100 (2005).
- Kendall, D. M. *et al.* Effects of exenatide (exendin-4) on glycemic control over 30 weeks in patients with type 2 diabetes treated with metformin and a sulfonylurea. *Diabetes Care* **28**, 1083–1091 (2005).
- Drucker, D. J. Biologic actions and therapeutic potential of the proglucagon-derived peptides. *Nature Clin. Pract. Endocrinol. Metab.* **1**, 22–31 (2005).
- Dakin, C. L. *et al.* Peripheral oxyntomodulin reduces food intake and body weight gain in rats. *Endocrinology* **145**, 2687–2695 (2004).
- Cohen, M. A. *et al.* Oxyntomodulin suppresses appetite and reduces food intake in humans. *J. Clin. Endocrinol. Metab.* **88**, 4696–4701 (2003).
- Baggio, L. L., Huang, Q., Brown, T. J. & Drucker, D. J. Oxyntomodulin and glucagon-like peptide-1 differentially regulate murine food intake and energy expenditure. *Gastroenterology* **127**, 546–558 (2004).
- Dakin, C. L. *et al.* Oxyntomodulin inhibits food intake in the rat. *Endocrinology* **142**, 4244–4250 (2001).
- Dakin, C. L. *et al.* Repeated ICV administration of oxyntomodulin causes a greater

- reduction in body weight gain than in pair-fed rats. *Am. J. Physiol. Endocrinol. Metab.* **283**, E1173–E1177 (2002).
73. Wynne, K. *et al.* Oxyntomodulin increases energy expenditure in addition to decreasing energy intake in overweight and obese humans: a randomised controlled trial. *Int. J. Obes. (Lond.)* advance online publication 18 April 2006 (doi:10.1038/sj.ijo.0803344).
 74. Martin, N. M. *et al.* Pre-obese and obese agouti mice are sensitive to the anorectic effects of peptide YY_{3–36} but resistant to ghrelin. *Int. J. Obes. Relat. Metab. Disord.* **28**, 886–893 (2004).
 75. Acuna-Goycolea, C. & van den Pol, A. N. Peptide YY_{3–36} inhibits both anorexigenic proopiomelanocortin and orexigenic neuropeptide Y neurons: implications for hypothalamic regulation of energy homeostasis. *J. Neurosci.* **25**, 10510–10519 (2005).
 76. Cone, R. D. Anatomy and regulation of the central melanocortin system. *Nature Neurosci.* **8**, 571–578 (2005).
 77. Abbott, C. R. *et al.* The inhibitory effects of peripheral administration of peptide YY_{3–36} and glucagon-like peptide-1 on food intake are attenuated by ablation of the vagal-brainstem-hypothalamic pathway. *Brain Res.* **1044**, 127–131 (2005).
 78. Browning, K. N. & Travagli, R. A. Neuropeptide Y and peptide YY inhibit excitatory synaptic transmission in the rat dorsal motor nucleus of the vagus. *J. Physiol. (Lond.)* **549**, 775–785 (2003).
 79. Chen, C. H. & Rogers, R. C. Peptide YY and the Y2 agonist PYY-(13–36) inhibit neurons of the dorsal motor nucleus of the vagus. *Am. J. Physiol.* **273**, R213–R218 (1997).
 80. Koda, S. *et al.* The role of the vagal nerve in peripheral PYY_{3–36}-induced feeding reduction in rats. *Endocrinology* **146**, 2369–2375 (2005).
 81. Date, Y. *et al.* The role of the gastric afferent vagal nerve in ghrelin-induced feeding and growth hormone secretion in rats. *Gastroenterology* **123**, 1120–1128 (2002).
 82. le Roux, C. W. *et al.* Ghrelin does not stimulate food intake in patients with surgical procedures involving vagotomy. *J. Clin. Endocrinol. Metab.* **90**, 4521–4524 (2005).
 83. Kuvshinov, B. W., Rudnicki, M. & McFadden, D. W. The effect of SMS 201–995 on meal and CCK-stimulated peptide YY release. *J. Surg. Res.* **50**, 425–429 (1991).
 84. Flanagan, D. E. *et al.* The influence of insulin on circulating ghrelin. *Am. J. Physiol. Endocrinol. Metab.* **284**, E313–E316 (2003).
 85. Ueno, M., Carvalheira, J. B., Oliveira, R. L., Velloso, L. A. & Saad, M. J. Circulating ghrelin concentrations are lowered by intracerebroventricular insulin. *Diabetologia* **49**, 2449–2452 (2006).
 86. Arafat, A. M. *et al.* Glucagon suppression of ghrelin secretion is exerted at hypothalamus-pituitary level. *J. Clin. Endocrinol. Metab.* **91**, 3528–3533 (2006).
 87. Neary, N. M. *et al.* Peptide YY_{3–36} and glucagon-like peptide-1_{7–36} inhibit food intake additively. *Endocrinology* **146**, 5120–5127 (2005).
 88. Talsania, T., Anini, Y., Siu, S., Drucker, D. J. & Brubaker, P. L. Peripheral exendin-4 and peptide YY_{3–36} synergistically reduce food intake through different mechanisms in mice. *Endocrinology* **146**, 3748–3756 (2005).
 89. Besterman, H. S. *et al.* Gut hormones in tropical malabsorption. *Br. Med. J.* **ii**, 1252–1255 (1979).
 90. Besterman, H. S. *et al.* Gut hormones in inflammatory bowel disease. *Scand. J. Gastroenterol.* **18**, 845–852 (1983).
 91. Adrian, T. E. *et al.* Peptide YY abnormalities in gastrointestinal diseases. *Gastroenterology* **90**, 379–384 (1986).
 92. Allen, J. M. *et al.* Effect of partial ileal bypass on the gut hormone responses to food in man. *Digestion* **28**, 191–196 (1983).
 93. Borg, C. M. *et al.* Progressive rise in gut hormone levels after Roux-en-Y gastric bypass suggests gut adaptation and explains altered satiety. *Br. J. Surg.* **93**, 210–215 (2006).
 94. Chan, J. L., Mun, E. C., Stoyneva, V., Mantzoros, C. S. & Goldfine, A. B. Peptide YY levels are elevated after gastric bypass surgery. *Obesity (Silver Spring)* **14**, 194–198 (2006).
 95. le Roux, C. W. *et al.* Gut hormone profiles following bariatric surgery favor an anorectic state, facilitate weight loss, and improve metabolic parameters. *Ann. Surg.* **243**, 108–114 (2006).

Acknowledgements We thank W. S. Dhillon and N. M. Martin for their assistance with the preparation of this manuscript. K.G.M. is supported by Biotechnology and Biological Sciences Research Council New Investigator Award.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to S.R.B. (s.bloom@imperial.ac.uk).

Inflammation and metabolic disorders

Gökhan S. Hotamisligil¹

Metabolic and immune systems are among the most fundamental requirements for survival. Many metabolic and immune response pathways or nutrient- and pathogen-sensing systems have been evolutionarily conserved throughout species. As a result, immune response and metabolic regulation are highly integrated and the proper function of each is dependent on the other. This interface can be viewed as a central homeostatic mechanism, dysfunction of which can lead to a cluster of chronic metabolic disorders, particularly obesity, type 2 diabetes and cardiovascular disease. Collectively, these diseases constitute the greatest current threat to global human health and welfare.

The incidence of obesity worldwide has increased drastically during recent decades. Consequently, obesity and associated disorders now constitute a serious threat to the current and future health of all populations on Earth. The World Health Organization estimates that more than 1 billion adults worldwide are overweight, 300 million of whom are clinically obese — defined as having a body mass index equal to or greater than 30 kg m⁻² (ref. 1). Particularly alarming is the equally marked increase in obesity among children². Obesity is associated with an array of additional health problems, including increased risk of insulin resistance, type 2 diabetes, fatty liver disease, atherosclerosis, degenerative disorders including dementia, airway disease and some cancers³ (Fig. 1). This cluster of pathologies has also started to emerge in children at young ages, a phenomenon that was inconceivable only a few decades ago.

This article focuses on obesity and type 2 diabetes, illustrating the links between nutrient- and pathogen-sensing pathways, and the interfacing of metabolic and inflammatory responses through these links as the mechanistic core of chronic and common metabolic diseases.

Key conceptual considerations

During the past decade, it became clear that inflammation is a key feature of obesity and type 2 diabetes⁴. Before delving into the mechanisms underlying metabolic dysfunction and the connections to inflammation, insulin action and metabolic disease clusters, it would be helpful to explore these terms and perhaps introduce a few crucial concepts that might differ from the classical context in which these terms have already been used and interpreted. Without venturing into the complexities of nomenclature, it would be useful to mention, at the onset of this discussion, that both 'inflammation' and 'metabolic syndrome' need to be redefined. This is essential to exploring the question of causality between the state of inflammation and the components that make up the cluster of metabolic pathologies (traditionally referred to as metabolic syndrome increasing metabolic disease risk)³ and therefore to developing effective mechanistic models and, ultimately, therapeutic strategies. For the purposes of this review, I occasionally refer to metabolic syndrome as a cluster of chronic and complex diseases all of which feature metabolic deterioration as at least one component.

With regard to inflammation, the traditional features of this state do not apply to the diseases in question. In the classic literature, inflammation is described as the principal response of the body invoked to deal with injuries, the hallmarks of which include swelling, redness, pain and fever (tumor, rubor, dolor and calor)⁵. This often short-term adaptive

response is a crucial component of tissue repair and involves integration of many complex signals in distinct cells and organs. However, the long-term consequences of prolonged inflammation are often not beneficial. This certainly seems to be the case in metabolic diseases. Although many of the same mediators are involved in obesity and diabetes, few, if any, of the classic features of inflammation have been observed. Therefore, it would be useful to set out a distinct form of injury response or subclass of inflammation — sometimes referred to as 'low-grade' or 'chronic' — or to describe an altogether separate state with a new term, perhaps 'metaflammation' (metabolically triggered inflammation). This condition is principally triggered by nutrients and metabolic surplus, and engages a similar set of molecules and signalling pathways to those involved in classical inflammation.

Evolutionary perspectives

Why are metabolic diseases so common and why are they so crucially linked to inflammatory processes? Perhaps we can gain insight through the fundamental biological design of an organism. Among the most critical processes to species survival are the ability to withstand starvation and the capacity to mount an effective immune response to pathogens. The former selects for energy efficiency and favours the storage of excess calories when access to food is intermittent. However, in the presence of a continuous nutritional surplus, this once advantageous metabolic state could set the stage for excess adiposity and its associated

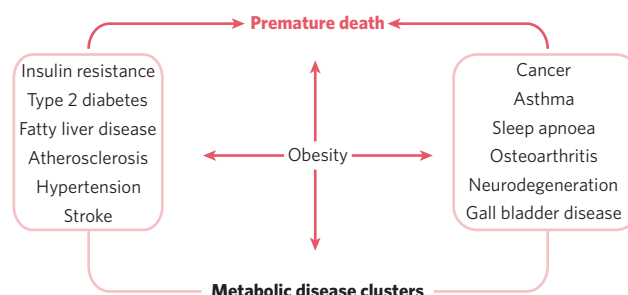


Figure 1 | Clustering of metabolic diseases. Obesity is considered to be a central feature that increases the risk for a vast array of diseases, with significant morbidity and mortality. In general, the mechanistic basis of the link between obesity and the diseases listed on the right is poorly understood compared with that of those listed on the left.

¹Department of Genetics & Complex Diseases, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.

problems. The ability to fight off an infection has also led to selection of strong immune responses, particularly after massive population declines during periods of infectious disease epidemics and pandemics^{6,7}. The combination of these traits is likely to have given rise to a biological organization that is highly capable of processing and storing energy and is also equipped with a powerful, and perhaps at times overly sensitive, immune response.

There is also an intimate relationship between the immune and metabolic response systems that has many evolutionary underpinnings. First, the functional units that control key metabolic and immune functions in higher organisms have evolved from common ancestral structures. One such structure is the *Drosophila* fat body, which incorporates the mammalian homologues of the liver and the haematopoietic and immune systems^{8,9} (Fig. 2). Interestingly, this site is also recognized as the equivalent of mammalian adipose tissue, sharing similar developmental and functional pathways^{10,11}. The fly's fat body carries out a crucial function in sensing energy and nutrient availability, and coordinates the appropriate metabolic and survival responses⁸. It is also the site of coordination of pathogen responses with metabolic status. In higher organisms, the adipose tissue, liver and haematopoietic system have specialized into distinct functional units or organs. However, these organs have maintained their developmental heritage, which was shared in earlier organisms. Therefore, it is possible to imagine a situation in which common or overlapping pathways regulate both metabolic and immune functions through common key regulatory molecules and signalling systems. This might allow nutrients to act through pathogen-sensing systems such as Toll-like receptors (TLRs), giving rise to metabolically or nutritionally induced inflammatory responses^{6,8,12,13}.

It is interesting to note that both adipose tissue and the liver have an architectural organization in which metabolic cells (adipocytes or hepatocytes) are in close proximity to immune cells (Kupffer cells or macrophages) and both have immediate access to a vast network of blood vessels (Fig. 3). With this configuration, both tissues form a suitable environment for continuous and dynamic interactions between immune and metabolic responses and establish communications with other sites such as pancreatic islets and muscle (Fig. 3). In fact, this interface might contribute to the emerging importance of these two organs in the initiation and development of metabolic diseases, particularly in the context of obesity and type 2 diabetes^{4,14}.

A closely linked configuration and coordinated regulation of metabolic and immune responses is likely to be advantageous in certain conditions, because an organism would need to organize and redistribute its energy resources during the mounting of an immune or inflammatory response. In fact, the most primitive response systems integrate the pathogen- and nutrient-sensing pathways such that nutrients can induce immune responses and pathogens can evoke and regulate metabolic responses^{6,8,15}. In this setting, we can foresee potential benefits to the activation of inflammatory or stress responses to block major anabolic signalling pathways such as the insulin/insulin growth factor (IGF) pathway. Blocking insulin signalling would divert energy sources from synthetic pathways. In this case, one might speculate that almost all stress and inflammatory signalling needs to coordinate with insulin-receptor signalling, and that all of these stress- and immune-response pathways are potentially involved in the disturbance of insulin action. However, none of these systems have evolved and adapted to be beneficial in the presence of continuous nutrient surplus such as we are now experiencing.

Taking all the evidence together, it is safe to suggest that the link between inflammatory and metabolic signalling is a delicate balance. Although there are short-term compensatory and adaptive measures to keep this delicate balance in check, the outcome is often detrimental when one arm overwhelms the other in the long term. For example, sustained exposure to pathogens or pathogen-associated components can disrupt systemic metabolic function from flies to humans^{15,16}. Similarly, chronic disturbance of metabolic homeostasis, such as occurs in malnutrition or overnutrition, could lead to aberrant immune responses⁴. The current nutritional habits and lifestyles of most modern humans

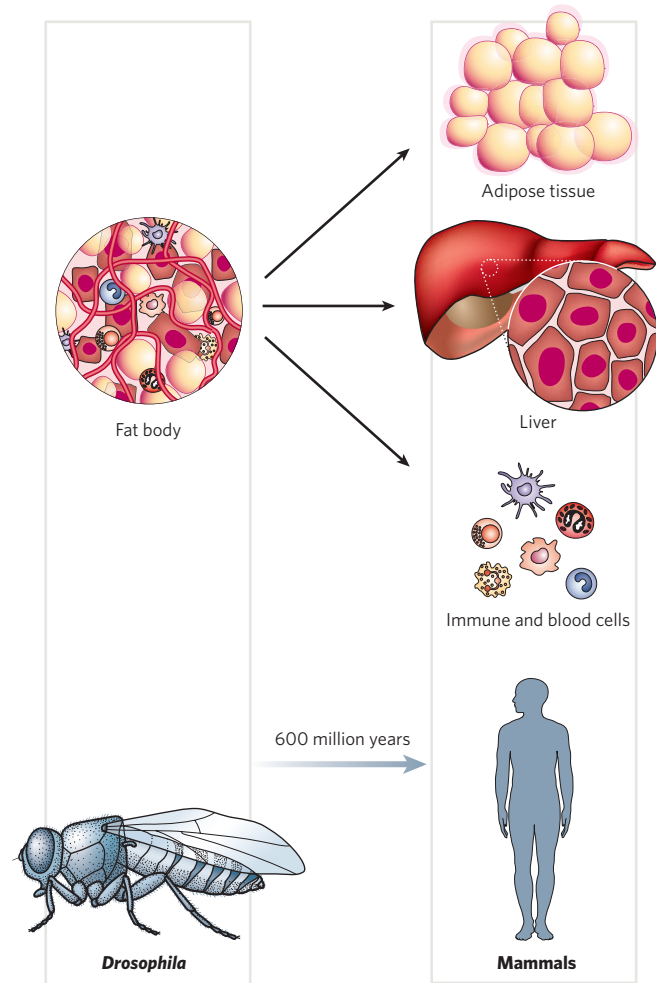


Figure 2 | Evolution of adipose tissue, the liver and the haematopoietic system into distinct organs in mammals. The adipose tissue, liver and haematopoietic system are all organized in one functional unit in *Drosophila melanogaster*, known as the 'fat body'. This developmental heritage may underlie the highly overlapping biological repertoire of these organs, their effects on metabolic and immune cells, and the close link between immune and metabolic response systems.

heavily favours metabolic overload with diminishing physical activity. Under such conditions, these historically advantageous traits and the juxtaposition of nutrient and pathogen responses have established the groundwork for chronic metabolic diseases to emerge and spread around the globe at an alarming pace. If this is the case, it is conceivable that prevention or treatment might be provided by targeting the same interface from which the metabolic disease clusters originate.

Obesity, inflammation and metabolic syndrome

Obesity, insulin resistance and type 2 diabetes are closely associated with chronic 'inflammation' characterized by abnormal cytokine production, increased acute-phase reactants and other mediators, and activation of a network of inflammatory signalling pathways⁴. Unequivocal experimental, epidemiological and clinical evidence produced during the past decade causally links inflammation, or the molecules and networks integral to inflammatory responses, to the development of these metabolic diseases and/or the complications that emerge from these pathologies, particularly in the context of obesity and type 2 diabetes^{4,14,17}.

The finding a little over a decade ago that tumour necrosis factor- α (TNF- α) is overexpressed in the adipose tissue of obese mice provided the first clear link between obesity, diabetes and chronic inflammation¹⁸.

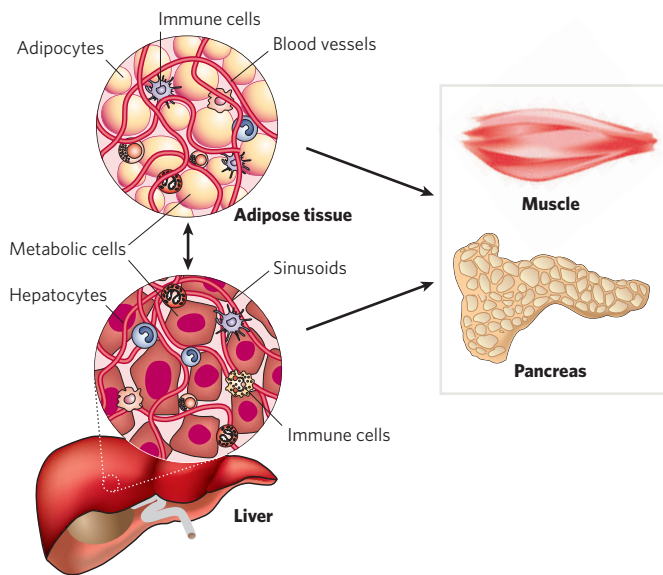


Figure 3 | Architectural organization and proximity of principal metabolic (adipocyte and hepatocyte) and immune (macrophages, Kupffer cells, lymphocytes and dendritic cells) cells in adipose tissue and liver. This configuration allows close interaction between these important cell groups and access to blood vessels for soluble mediators that communicate with other organs, including, but not limited to, skeletal muscle and the pancreas. This network is detrimental when inflammation is heightened, as in adipose tissue and/or the liver in individuals with obesity and/or type 2 diabetes.

TNF- α is a proinflammatory cytokine that activates various signal transduction cascades, including many of the pathways that are discussed below as critical inhibitors of insulin action. In obese mouse models a lack of TNF- α function results in improved insulin sensitivity and glucose homeostasis, confirming that this inflammatory response has a critical role in the regulation of insulin action in obesity^{19,20}. TNF- α is also overexpressed in the adipose and muscle tissues of obese humans, and when administered exogenously leads to insulin resistance^{21–24}. Whether TNF- α alone is involved in human insulin resistance has not yet been adequately addressed because of the limited power and scope of the early studies, which did not yield a detectable benefit of TNF- α neutralization²⁵. Interestingly, however, the widespread use of anti-TNF- α treatments in inflammatory diseases such as rheumatoid arthritis have produced clear secondary results supporting a role for TNF- α in systemic insulin sensitivity in humans^{26,27}.

It is clear from studies published during the past decade that, in addition to TNF- α , various other inflammatory mediators and cytokines are also overexpressed in adipose and other tissues in experimental mouse models of obesity and in humans⁴, and the measurable contribution of each inflammatory mediator is, at best, partial, even in experimental models. As cytokines and chemokines work in networks, the effect of individual molecules on metabolic function depends on their place in the hierarchy of the network; those more potent and proximal, such as TNF- α , have greater effects. The combinatorial and additive action of cytokines in metabolic homeostasis is a crucial but poorly addressed area.

Adipose tissue and immune response

The inflammatory response that emerges in the presence of obesity seems to be triggered by and to reside predominantly in adipose tissue, although other metabolically critical sites, particularly the liver, might also be involved during the course of the disease^{4,14}. Recent studies have documented the unusual properties of adipocytes and centrally placed adipose tissue as a crucial site in the generation of inflammatory responses and mediators. In addition to the inherent properties of fat cells in energy management and metabolic homeostasis, adipose tissue

serves as a key site for the interaction of adipocytes with other effectors of the immune system (Fig. 3).

There are also striking commonalities between adipocytes and a diverse set of immune cells (including T cells, macrophages and dendritic cells). These features are extensive and range from complement activation and production of inflammatory mediators to pathogen sensing and phagocytic properties^{4,14}. Thus, a potential window through which to explore the link between metabolism and inflammation might lie at the functional interface of cells of primarily immune or metabolic nature, such as adipocytes and macrophages, and the shared response systems. Although adipose tissue is presented here as an example of functional organization and structural proximity of metabolic and immune cells, a similar argument could be made for other metabolically critical organs, particularly the liver (Fig. 3).

An important feature of inflammation is infiltration of inflamed tissues by immune cells such as neutrophils, eosinophils and macrophages. Macrophage infiltration of adipose tissue has recently been described in obese conditions in both mice and humans^{28,29}. It has been suggested that expanding adipocytes or neighboring pre-adipocytes might begin to produce chemotactic signals leading to macrophage recruitment⁴. Careful examination of the morphology of adipose tissue during obesity has illustrated the convergence of macrophages on necrotic adipocytes and suggested that their presence in adipose tissue might be predominantly for clearance purposes³⁰. The death of adipocytes might be a primary event enhanced by obesity, or secondary to inflammation and macrophage infiltration. Thus, it might reflect an attempt to limit the expansion of these cells.

Although it is likely that macrophage infiltration into adipose tissue makes some contribution to the emergence and maintenance of obesity-induced inflammatory responses, the extent of this and the functional involvement of macrophages in systemic metabolism are unclear. Genetic manipulation of IKK- β (inhibitor of nuclear factor- κ B (NF- κ B) kinase- β) signalling in myeloid lineage can affect systemic metabolic regulation³¹, supporting the involvement of macrophages and/or neutrophils in insulin action. The potential participation of other immune cells such as lymphocytes remains unclear. Interestingly, disruption of insulin action in the myeloid cells does not seem to alter systemic glucose or lipid metabolism, although this question has only been addressed in atherosclerotic models^{32,33}. Additional pathways at the interface between macrophage and adipocyte function in metabolic diseases are further discussed in the context of lipid signals below. However, the understanding of interactions between immune and metabolic cells is incomplete.

Lipids in inflammation and insulin resistance

Metabolic, inflammatory and innate immune processes are also coordinately regulated by lipids³⁴. Several transcription factors, particularly those in the peroxisome-proliferator activated receptor (PPAR) and liver X receptor (LXR) families, seem to be crucial for modulating the intersection of these pathways. Activation of these transcription factors inhibits the expression of several genes involved in inflammatory response in macrophages and adipocytes^{35,36}.

Ligands to all three PPAR family members suppress production of proinflammatory cytokines, mostly through suppression of NF- κ B^{35,36}. Notably, unliganded PPAR- δ seems to have inflammatory functions, mediated at least in part through its association with the transcriptional repressor BCL-6 (ref. 37). Like the PPARs, LXR suppresses production of inflammatory mediators in a manner reciprocal to its regulation of lipid metabolism³⁸. Signalling from TLRs inhibits LXR activity in macrophages, causing enhanced cholesterol accumulation in macrophages and accounting, at least in part, for the pro-atherogenic effects of infection³⁹. Most intriguingly, TLRs are present in adipocytes and can be directly activated by nutrients, particularly fatty acids^{12,13,40}. Finally, there are critical interactions between different classes of nuclear hormone receptor that might represent a critical component of the anti-inflammatory actions of these molecules^{41,42}.

Another group of molecules that coordinates lipid responses in adipocytes and macrophages and that is linked to the pathways described

above is the lipid chaperone proteins. Animals lacking the adipocyte/macrophage fatty-acid-binding proteins (FABPs) aP2 and MAL1 exhibit a phenotype akin to that of mice and humans treated with PPAR γ ligands, indicating that FABPs and PPARs might act on similar pathways in controlling the biological effects of lipids^{43,44}. Inflammatory pathways, including the activation of c-Jun amino-terminal kinase (JNK) and IKK, and insulin action in target cells are also strongly influenced by these FABPs. Mice lacking aP2 and MAL1 are protected against almost every aspect of metabolic syndrome, including visceral obesity, insulin resistance, hepatosteatosis and atherosclerosis^{43,45,46}. Interestingly, a rare loss-of-function genetic variant at the aP2 locus produces a very similar protective phenotype in humans⁴⁷. Most recently, resistance to asthma was also demonstrated in aP2-deficient mice, establishing the first potential link between metabolic regulation and airway inflammation⁴⁸. Although our understanding of the precise mechanisms that underlie FABP action is incomplete, it is clear that their absence is both anti-inflammatory and protective of metabolic function during nutrient overload⁴⁹. FABPs also modulate lipid composition and fluxes, a critical feature of their biological function. In fact, it seems that location in the body is a crucial factor for determining whether lipids promote or suppress inflammation and insulin resistance. Even elevated plasma lipid levels might not be inflammatory alone, but instead might be correlated with the extent of stimulation of inflammatory pathways, because the hyperlipidaemic state is indicative of redistribution of lipids from adipose tissue to muscles and the liver.

Inflammatory signalling and insulin action

How do inflammatory signals disrupt insulin action and mediate insulin resistance in obesity? The insulin signalling pathway is not presented here in detail because of space limitations, and readers are referred to excellent recent reviews^{50,51}. However, to summarize briefly, insulin and IGF receptors belong to a receptor tyrosine kinase family and, unlike other receptors in this family, use docking proteins to mediate their signalling. Among a large number of intracellular substrates used by these receptors, six belong to the family of insulin receptor substrate (IRS) proteins (IRS-1–6)^{50,51}. Insulin stimulates tyrosine phosphorylation of IRS proteins, which is a crucial event in mediating insulin action. This step in insulin-receptor signalling is defective in most cases of systemic insulin resistance, both in experimental models and in humans⁴. TNF- α also targets this element of insulin-receptor signalling through inhibitory serine phosphorylation of IRS-1 (ref. 52). As this post-translational modification of IRS-1 through serine phosphorylation has also been detected in insulin-resistant cells and tissues in mice and humans, elucidation of the mechanisms that lead to phosphorylation of IRS-1 and other substrates has become a major focus in many laboratories in recent years.

It has now been established that IRS-1 is phosphorylated at serine residues by various kinases that interfere with the ability of this protein to engage in insulin-receptor signalling and result in alterations in insulin action^{50,53,54} (Fig. 4). In addition, suppressor of cytokine signalling (SOCS) proteins seem to inhibit insulin action at the level of insulin-receptor substrates, although through a different mechanism^{55–57}.

Among these IRS-modifying enzymes, mounting evidence indicates that activation of JNK, IKK and conventional protein kinase C (PKC) is central to mediating insulin resistance in response to various stresses that occur in obesity and other conditions of insulin resistance (Fig. 4). They have all been reported to be able to inhibit insulin action by serine phosphorylation of IRS-1 (refs 53, 58, 59), although the activity of IKK in this regard has not yet been well established under physiological conditions. IRS-1 serine phosphorylation disrupts insulin-receptor signalling through several distinct mechanisms and blocks insulin action^{60,61}. These kinases also exert powerful effects on gene expression, including promoting further inflammatory gene expression through activation of activator protein-1 (AP-1) complexes and NF- κ B⁶². However, this aspect has not yet been thoroughly explored in metabolic homeostasis.

Most importantly, in obesity there is a striking increase in JNK activity in critical metabolic sites such as adipose and liver tissues⁶³ as well as in the hypothalamus⁶⁴. JNK is activated upon exposure to cytokines

such as TNF- α , as well as by free fatty acids and internal cues such as endoplasmic reticulum stress, all of which might underlie the obesity-induced activity^{4,53,65}. Genetic JNK1-deficiency protects mice from obesity-induced JNK activation, IRS-1 serine phosphorylation, and, consequently, insulin resistance, fatty liver and diabetes^{63,66}. Recent studies examining the functional interactions between JNK isoforms revealed that JNK2 also participates in metabolic regulation but that this function is normally masked owing to compensation by JNK1 (ref. 66). JNK2 activity has also been implicated in the pathogenesis of atherosclerosis⁶⁷. It is not yet clear whether these pathological states are exclusive to the aberrant action of individual JNK isoforms or a function of total JNK activity in specific target tissues. Regardless of this uncertainty, interventions to block JNK activity in established models of obesity and diabetes improved systemic glucose homeostasis and insulin sensitivity, as well as atherosclerosis, suggesting that JNK inhibition might be a promising therapeutic avenue for diabetes^{67–69}.

Another inflammatory kinase that is critical in the development of insulin resistance and metabolic dysfunction is IKK- β . Mice that are heterozygous for a null mutation in IKK- β are partly protected from obesity-induced insulin resistance, and inhibition of IKK- β by administration of high-dose salicylates improves insulin action in experimental models and humans^{70,71}. Studies involving tissue- or cell-type specific modulation of IKK- β reveal that its activity in the liver affects systemic metabolism; however, its role in muscles is probably less important in terms of metabolic regulation⁷². Similarly to JNK, experimental activation of this kinase in the liver seems to be sufficient to generate systemic insulin resistance⁷². Interestingly, myeloid-specific deletion of IKK- β results in partial protection against obesity- or lipopolysaccharide-induced insulin resistance, providing clear evidence that IKK- β activity in myeloid cells can participate in the regulation of systemic metabolic homeostasis³¹.

Protein kinase C is also important in interaction between the inflammatory and metabolic pathways, and in particular has been implicated as a kinase downstream of lipid signals^{34,73,74}. Fatty acid metabolites such as

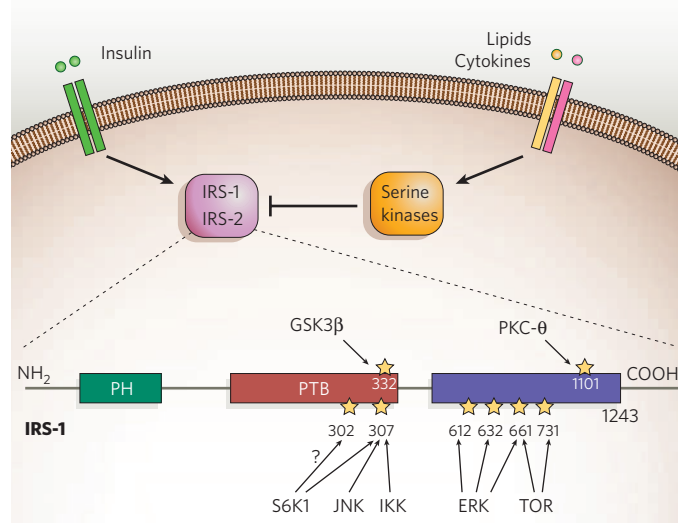


Figure 4 | Insulin-receptor signalling interfaces with inflammatory signalling at the level of insulin-receptor substrates through activation of serine kinases. These kinases respond to both lipids and cytokines. Both IRS-1 and IRS-2 contain a number of serine residues (indicated by white stars) that are targeted by various kinases, which have been more extensively mapped for IRS-1 (ref. 50). All of the molecules listed that regulate insulin action in response to inflammatory signals converge on IRS-1. Many serine phosphorylation sites also seem to exist in IRS-2 but have not yet been characterized in detail. ERK, extracellular regulated mitogen-activated protein kinase; GSK3 β , glycogen synthase kinase 3 β ; PH, pleckstrin homology; PTB, phosphotyrosine binding; S6K1, ribosomal protein S6 kinase polymerase 1; TOR, target of rapamycin.

fatty acyl coenzyme As and diacylglycerides can activate PKC- θ in muscles or PKC- δ in the liver and inhibit insulin action^{34,74}. Mice deficient in PKC- θ are protected against fatty-acid-induced insulin resistance, confirming the contribution of this kinase to metabolic regulation *in vivo*⁷³. At the mechanistic level, PKC- θ is known to activate IKK and might contribute to insulin resistance through this pathway. Interestingly, interactions also occur between PKC, JNK and IKK, although these have not been tested in the context of metabolic homeostasis⁷⁵. Characterization of the isoforms of these kinases involved in disrupting insulin action among different species, especially humans, is a critical issue that has so far been addressed only in part.

Endoplasmic reticulum stress in obesity and type 2 diabetes

A fundamental issue that is not yet well understood is why inflammation develops in obesity. What stresses initially cause the activation of inflammatory pathways? Are these signalling pathways emerging from a common mechanistic platform and integrating with each other? Recent data from experimental models indicate that endoplasmic reticulum (ER) stress is critical to the initiation and integration of pathways of inflammation and insulin action in obesity and type 2 diabetes.

The ER is a vast network of membranes in which all the secretory and membrane proteins are assembled into their secondary and tertiary structures. Proper folding, maturation, storage and transport of these proteins take place in this organelle. Unfolded or misfolded proteins are detected, removed from the ER and degraded by the 26S proteasome system^{76,77}. Accumulation of unfolded proteins, energy and nutrient fluctuations, hypoxia, toxins, viral infections and increased demand on the synthetic machinery give rise to perturbations in the ER lumen and create stress. Under these conditions, the ER activates a complex response

system known as the unfolded protein response (UPR) to restore the functional integrity of the organelle^{76,77}. The principal branches of UPR signalling are mediated through three molecules: inositol-requiring enzyme 1 (IRE-1), PKR-like endoplasmic-reticulum kinase (PERK) and activating transcription factor 6 (ATF6)^{78,79}.

Notably, the two principal inflammatory pathways that disrupt insulin action, JNK-AP-1 and IKK-NF- κ B, are linked to IRE-1 and PERK activity during ER stress⁸⁰⁻⁸². IRE-1 is linked to activation of JNK through a pathway involving TNF-receptor-associated factor 2 (ref. 80). Activation of both IRE-1 and PERK is also linked to the IKK-NF- κ B pathway, although through distinct mechanisms. Whereas IRE-1 interacts with IKK- β through TNF-receptor-activated factor 2 (TRAF2), PERK activation leads to degradation of I κ B and therefore facilitates the activity of NF- κ B^{81,82}.

So, we can postulate that the ER might be a site for the sensing of metabolic stress and the translation of that stress into inflammatory signalling and responses. In fact, the ER could be considered an essential and ancient site of integration between nutrient and pathogen responses as it is very sensitive to glucose and energy availability, lipids, pathogens and pathogen-associated components. In addition to increased demand on the synthetic capacity of several organs such as the liver and adipose tissue, increased adiposity in obesity produces an environment that further challenges ER function and capacity owing to architectural constraints that limit ER expansion as well as altered energy and nutrient availability. So, obesity provides many conditions that could lead to ER stress.

Work in our laboratory has recently demonstrated that in both dietary and genetic obesity ER stress is increased in adipose and liver tissues⁶⁵. In cellular systems, the induction of ER stress leads to insulin resistance, at least in part through IRE-1-dependent activation of JNK. Modulation of ER-folding capacity through gain- and loss-of-function studies with X-box binding protein (XBP-1) showed a close link between ER function and insulin action *in vitro* and *in vivo*⁶⁵. These initial observations were followed by independent studies that linked ER function to insulin sensitivity^{83,84}. Together, these results indicate that ER stress has an important role in mediating insulin resistance in obesity in animal models and that increasing cellular folding capacity might be a promising therapeutic approach, if this concept is applicable to humans. Indeed, recent studies have demonstrated that orally active small-molecule chemical chaperones are extremely effective in alleviating obesity-induced ER stress and JNK activation, and in treating insulin resistance and type 2 diabetes in mice⁸⁵.

ER stress also provides several links with the emergence of inflammatory responses. First, and as stated above, ER stress leads to activation of both JNK and IKK⁸⁰⁻⁸². Second, inflammatory mediators can trigger ER stress and lead to propagation of general stress responses⁸⁶. Third, ER stress leads to activation of CREB-H, which may have an important role in inflammatory and acute-phase responses in the liver⁸⁶. Fourth, the ER is a major source of reactive oxygen species (ROS), and, consequently, oxidative stress in all cells^{87,88}. Oxidative stress is emerging as a feature of obesity and an important factor in the development of insulin resistance in obesity⁸⁹⁻⁹¹.

In metabolic disease, the molecular outcomes of oxidative stress have been most clearly linked to diabetic complications through endothelial cell dysfunction⁹². However, recent evidence indicates that oxidative stress and mitochondrial dysfunction also have important roles in type 2 diabetes^{89-91,93}. These pathways are also coupled to activation of inflammatory pathways and insulin resistance in adipocytes and muscle cells, and impaired insulin secretion in pancreatic β -cells⁸⁹⁻⁹¹. Both the NF- κ B and JNK pathways can be activated under conditions of oxidative stress, and this may be important for the ability of ROS to mediate insulin resistance. Taken together, the effector arms of ER stress and the associated stress responses are tightly linked to inflammatory pathways at many levels that are crucial for insulin action and metabolic homeostasis.

Space limitations preclude detailing the implications of the mechanisms discussed here for islet function and survival. However, all of the pathways that act at the interface of metabolic and inflammatory

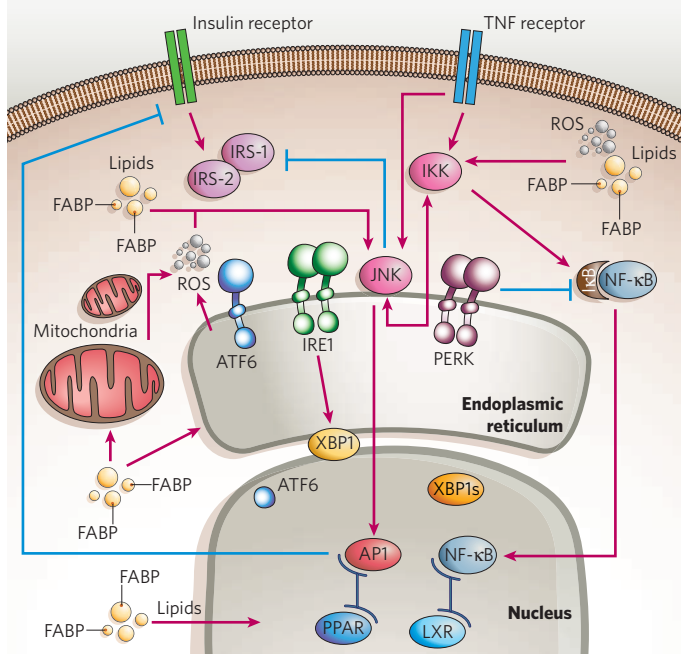


Figure 5 | Molecular pathways integrating stress and inflammatory responses with insulin action. IRS-1 and 2 are crucial signalling molecules in insulin action. Activation of JNK by cytokine signalling, lipid products, ROS or through IRE1 during ER stress leads to serine phosphorylation of IRS-1 and 2, and consequently inhibits insulin signalling. Similar signals, including PERK, also activate IKK and inhibition of insulin action through a series of transcriptional events mediated by NF- κ B. JNK also regulates transcription through AP-1. Lipid-activated transcriptional events are mediated by nuclear hormone receptors PPAR and LXR. The biological activities of lipids are regulated by FABPs that function as chaperones. Mitochondria and the ER can both contribute to ROS production. ATF6 and XBP1 are critical regulators of ER function and its adaptive responses.

responses have critical direct and/or indirect effects on β -cells. This is particularly true in ER stress and related signalling pathways^{76,77}. Hence, these mechanisms have great potential to underscore the integrated pathophysiology of type 2 diabetes characterized by both insulin resistance and defective insulin secretion. Also not discussed is the action of insulin on the central nervous system (CNS) and consequences of CNS insulin resistance, not only in systemic metabolic homeostasis but also in neurodegeneration and dementia. It has even been proposed that defective insulin action in the nervous system be labelled type 3 diabetes, particularly in the context of Alzheimer's disease⁹⁴. It is interesting to note that the mitochondrial effects of the PPAR- γ coactivator-1 α (PGC-1 α) have an important role in oxidative stress in both type 2 diabetes and neurodegenerative disease^{93,95,96}.

The chicken or egg question

There is experimental evidence clearly demonstrating that, under defined conditions, inflammatory mediators 'alone' can trigger insulin resistance in cells, experimental models and humans in the absence of other triggering factors, such as obesity. This observation, on the surface, suggests that inflammation is proximal to metabolic deterioration. Although this concept holds true for selected conditions, such as severe infections, burns and trauma, it is not clear whether the same is applicable to obesity or other chronic metabolic disorders. Furthermore, it is also clear that metabolic dysfunction can be triggered by chronic excess of nutrients, such as lipids and glucose. However, in this case, these signals also simultaneously trigger inflammatory responses. In this setting of metabolic excess, it is more likely that metabolic signal(s) are the triggers for inflammatory responses, which then further disrupt metabolic function, leading to more stress and inflammation, and so on. Such vicious cycles could be described for all of the critical events and identified mechanisms leading to further metabolic deterioration (Fig. 5).

For example, at the molecular and cellular levels, nutrients such as lipids and cytokines can trigger inflammatory kinases, such as JNK and IKK, and ER stress. The presence of ER stress, regardless of the proximal signal, can also activate JNK and IKK, both of which then regulate the production of various cytokines, including TNF- α . Notably, exposure to inflammatory cytokines such as TNF- α can induce ER stress, and ER stress itself can cause an increase in the expression of TNF- α or perhaps more general inflammatory responses⁸². A similar situation could occur for ROS, which might emerge from mitochondria and/or the ER and activate JNK and IKK. This would cause more ER stress, block insulin action and produce more ROS, which would, in turn, produce broader inflammatory responses. Some of the pathways integrating inflammation with insulin action are detailed in Fig. 5. As this scheme shows, it is possible to enter the network from an inflammatory or metabolic gateway and still end up with the same vicious loop.

Another way to look into the chicken or egg question is to ask whether inflammation can simply be a function of inefficient nutrient clearance. If this is the case, as long as the apparatus dealing with nutrient clearance is operating effectively, there should not be inflammation related to metabolic signals. Although this is certainly a valid consideration, it would require several issues to be resolved. For example, a clearance-related theory would imply that, under physiological or homeostatic conditions, there are yet to be identified mechanisms that can prevent the engaging of inflammatory pathways during nutrient fluctuations. Furthermore, clearance must not only be efficient but must also take place in appropriate sites — for example, glucose by the muscle and fat by adipose tissue. Even if these mechanisms were present, excess nutrients could still trigger stress/inflammatory responses during their intracellular metabolism. Perhaps control at this level might involve the anti-inflammatory actions of glucocorticoids and their functional interactions with other nuclear hormone receptors responding to potential nutrient-derived signals^{41,42}. Nevertheless, in an integrated homeostatic model, long-term exposure to nutrient excess is highly unlikely to be managed simply by adjusting the ability to clear and enhance metabolism at target cells. Discovery of such sensing and control paradigms

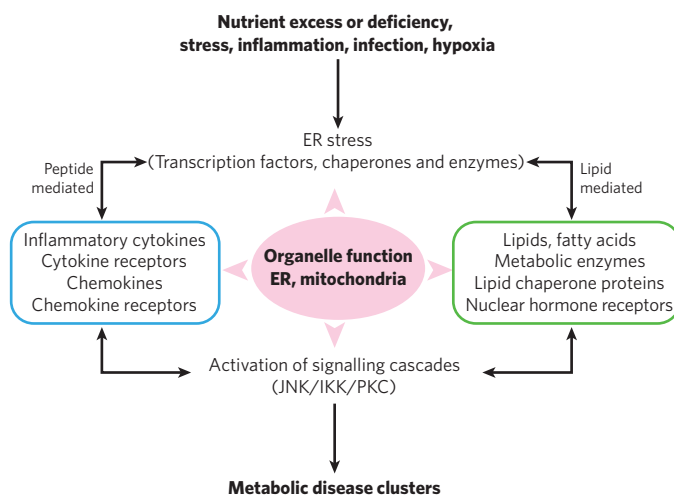


Figure 6 | Therapeutic targets at the interface between metabolic and inflammatory pathways. The pathways are divided into peptide- and lipid-mediated targets for practical purposes and do not represent an exhaustive list. Treating several loci involved in the disease process by targeting organelles such as the ER and mitochondria represents a new approach to treating metabolic diseases.

might nevertheless open the way for therapeutic approaches to enhance the capacity of the endogenous machinery so that stress/inflammatory responses are prevented during exposure to the same flow of metabolic signals.

Therapeutic opportunities

This article opened with the statement that metabolic and immune systems are among the most fundamental requirements for survival. How, then, can we manipulate the interface of such fundamental biological response pathways for therapeutic purposes without severe consequences to the organism? This is a formidable challenge but some promise exists and is emerging.

Some potential therapeutic avenues are summarized in Fig. 6. For practical purposes, these are divided into peptide- and lipid-mediated pathways. For peptides, the most obvious targets are cytokines, chemokines or their receptors. Although there have been some encouraging results in relation to this approach (for example anti-TNF or anti-CCR2 (chemokine (C-C motif) receptor 2) therapies), it is likely that the benefits of targeting a single cytokine or signalling receptor will be limited^{41,44}.

A more comprehensive approach would be to tackle a more central locus in the production of not single molecules but a network of responses. The best examples for this approach would be JNK and IKK pathways. Through inhibition of IKK, high-dose salicylates can improve glucose metabolism in both obese mice and diabetic humans^{70,71,97}. It is possible that inhibition of other inflammatory kinases might have similar effects. In the preclinical setting, genetic, biochemical and pharmacological targeting of inflammatory signalling pathways improves insulin action⁴. For example, the targeting of JNK using an inhibitory peptide, synthetic small-molecule inhibitors or RNA interference (RNAi)-based technologies has been shown to improve insulin action in various mouse models^{68,69}. There is also strong genetic evidence that the JNK pathway is relevant to human disease⁹⁸. However, developing effective orally active small-molecule inhibitors of JNK that are isoform selective has been challenging. This still constitutes a bottleneck in addressing the therapeutic potential of targeting JNK in humans and needs to be addressed.

Among the lipid-related pathways, the main example of therapeutic success is the thiazolidinediones⁹⁹. These PPAR γ ligands are insulin-sensitizing compounds, in use for humans, that function by regulating lipid metabolism and exhibiting anti-inflammatory effects. FABPs can also be targeted by small molecules to inhibit both adipocyte and

macrophage function, and to treat type 2 diabetes and atherosclerosis in mice^{43,46}. Although many other possible molecules could be listed, those that engage directly with lipids as ligands or signals and exhibit the dual activity on metabolic and inflammatory pathways are mentioned here.

A final and truly paradigm-shifting approach would be 'organelle therapy'. As mitochondrial defects and ER dysfunction are central to the activation of several important inflammatory pathways, chemical correction of their functional deficiency might result in new treatments to stop the vicious cycle between metabolic and inflammatory cascades and rescue insulin action and/or correct metabolic disorders. In such a case, several of the mechanisms underlying metabolic dysfunction could be treated simultaneously for superior efficacy and safety. The use of chemical chaperones in experimental models supports the feasibility of such an approach in metabolic disease, although it is not yet clear whether such concepts are applicable to human disease⁸⁵. Recently, the possibility of targeting mitochondrial dysfunction has also been suggested¹⁰⁰. In searching for new and effective therapeutics, it might be useful to use a systems-chemistry approach to modify integrated outcomes rather than targeting single molecules with the hope that the desired systematic effect might be generated. In other words, it is likely that creating a 'new homeostasis' will require the modification of more than one target. ■

- The World Health Report 2002 *Reducing Risks, Promoting Healthy Life* (World Health Organization, Geneva, 2002).
- Rocchini, A. P. Childhood obesity and a diabetes epidemic. *N. Engl. J. Med.* **346**, 854–855 (2002).
- Semenkovich, C. F. Insulin resistance and atherosclerosis. *J. Clin. Invest.* **116**, 1813–1822 (2006).
- Wellen, K. E. & Hotamisligil, G. S. Inflammation, stress, and diabetes. *J. Clin. Invest.* **115**, 1111–1119 (2005).
- Larsen, G. L. & Henson, P. M. Mediators of inflammation. *Annu. Rev. Immunol.* **1**, 335–359 (1983).
- Beutler, B. Innate immunity: an overview. *Mol. Immunol.* **40**, 845–859 (2004).
- Levin, B. R., Lipsitch, M. & Bonhoeffer, S. Population biology, evolution, and infectious disease: convergence and synthesis. *Science* **283**, 806–809 (1999).
- Sondergaard, L. Homology between the mammalian liver and the *Drosophila* fat body. *Trends Genet.* **9**, 193 (1993).
- Leclerc, V. & Reichhart, J. M. The immune response of *Drosophila melanogaster*. *Immunol. Rev.* **198**, 59–71 (2004).
- Tong, Q. *et al.* Function of GATA transcription factors in preadipocyte-adipocyte transition. *Science* **290**, 134–138 (2000).
- Rusten, T. E. *et al.* Programmed autophagy in the *Drosophila* fat body is induced by ecdysone through regulation of the PI3K pathway. *Dev. Cell* **7**, 179–192 (2004).
- Song, M. J., Kim, K. H., Yoon, J. M. & Kim, J. B. Activation of Toll-like receptor 4 is associated with insulin resistance in adipocytes. *Biochem. Biophys. Res. Commun.* **346**, 739–745 (2006).
- Shi, H. *et al.* TLR4 links innate immunity and fatty acid-induced insulin resistance. *J. Clin. Invest.* **116**, 3015–3025 (2006).
- Shoelson, S. E., Lee, J. & Goldfine, A. B. Inflammation and insulin resistance. *J. Clin. Invest.* **116**, 1793–1801 (2006).
- Dionne, M., Pham, L. N., Shirasu-Hiza, M. & Schneider, D. S. Akt and foxo dysregulation contribute to infection-induced wasting in *Drosophila*. *Curr. Biol.* **16**, 1977–1985 (2006).
- Charlton, M. R., Pockros, P. J. & Harrison, S. A. Impact of obesity on treatment of chronic hepatitis C. *Hepatology* **43**, 1177–1186 (2006).
- Dandona, P., Aljada, A. & Bandyopadhyay, A. Inflammation: the link between insulin resistance, obesity and diabetes. *Trends Immunol.* **25**, 4–7 (2004).
- Hotamisligil, G. S., Shargill, N. S. & Spiegelman, B. M. Adipose expression of tumor necrosis factor- α : direct role in obesity-linked insulin resistance. *Science* **259**, 87–91 (1993).
- Uysal, K. T., Wiesbrock, S. M., Marino, M. W. & Hotamisligil, G. S. Protection from obesity-induced insulin resistance in mice lacking TNF- α function. *Nature* **389**, 610–614 (1997).
- Ventre, J. *et al.* Targeted disruption of the tumor necrosis factor- α gene — metabolic consequences in obese and nonobese mice. *Diabetes* **46**, 1526–1531 (1997).
- Hotamisligil, G. S., Arner, P., Caro, J. F., Atkinson, R. L. & Spiegelman, B. M. Increased adipose tissue expression of tumor necrosis factor- α in human obesity and insulin resistance. *J. Clin. Invest.* **95**, 2409–2415 (1995).
- Kern, P. A. *et al.* The expression of tumor necrosis factor in human adipose tissue. Regulation by obesity, weight loss, and relationship to lipoprotein lipase. *J. Clin. Invest.* **95**, 2111–2119 (1995).
- Saghiadeh, M., Ong, J. M., Garvey, W. T., Henry, R. R. & Kern, P. A. The expression of TNF α by human muscle: relationship to insulin resistance. *J. Clin. Invest.* **97**, 1111–1116 (1996).
- Krogh-Madsen, R., Plomgaard, P., Moller, K., Mittendorfer, B. & Pedersen, B. K. Influence of TNF- α and IL-6 infusions on insulin sensitivity and expression of IL-18 in humans. *Am. J. Physiol. Endocrinol. Metab.* **291**, E108–E114 (2006).
- Ofei, F., Hurel, S., Newkirk, J., Sopwith, M. & Taylor, K. Effects of an engineered human anti-TNF- α antibody (CDP571) on insulin sensitivity and glycemic control in patients with NIDDM. *Diabetes* **45**, 881–885 (1996).
- Gonzalez-Gay, M. A. *et al.* Anti-tumor necrosis factor- α blockade improves insulin resistance in patients with rheumatoid arthritis. *Clin. Exp. Rheumatol.* **24**, 83–86 (2006).
- Kiortsis, D. N., Mavridis, A. K., Vasakos, S., Nikas, S. N. & Drosos, A. A. Effects of infliximab treatment on insulin resistance in patients with rheumatoid arthritis and ankylosing spondylitis. *Ann. Rheum. Dis.* **64**, 765–766 (2005).
- Weisberg, S. P. *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *J. Clin. Invest.* **112**, 1796–1808 (2003).
- Xu, H. *et al.* Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J. Clin. Invest.* **112**, 1821–1830 (2003).
- Cinti, S. *et al.* Adipocyte death defines macrophage localization and function in adipose tissue of obese mice and humans. *J. Lipid Res.* **46**, 2347–2355 (2005).
- Arkan, M. C. *et al.* IKK- β links inflammation to obesity-induced insulin resistance. *Nature Med.* **11**, 191–198 (2005).
- Baumgartl, J. *et al.* Myeloid lineage cell-restricted insulin resistance protects apolipoprotein E-deficient mice against atherosclerosis. *Cell Metab.* **3**, 247–256 (2006).
- Han, S. *et al.* Macrophage insulin receptor deficiency increases ER stress-induced apoptosis and necrotic core formation in advanced atherosclerotic lesions. *Cell Metab.* **3**, 257–266 (2006).
- Yu, C. *et al.* Mechanism by which fatty acids inhibit insulin activation of insulin receptor substrate-1 (IRS-1)-associated phosphatidylinositol 3-kinase activity in muscle. *J. Biol. Chem.* **277**, 50230–50236 (2002).
- Chawla, A., Repa, J. J., Evans, R. M. & Mangelsdorf, D. J. Nuclear receptors and lipid physiology: opening the X-files. *Science* **294**, 1866–1870 (2001).
- Glass, C. K. & Ogawa, S. Combinatorial roles of nuclear receptors in inflammation and immunity. *Nature Rev. Immunol.* **6**, 44–55 (2006).
- Lee, C. H., Olson, P. & Evans, R. M. Mini-review: lipid metabolism, metabolic diseases, and peroxisome proliferator-activated receptors. *Endocrinology* **144**, 2201–2207 (2003).
- Joseph, S. B., Castrillo, A., Laffitte, B. A., Mangelsdorf, D. J. & Tontonoz, P. Reciprocal regulation of inflammation and lipid metabolism by liver X receptors. *Nature Med.* **9**, 213–219 (2003).
- Castrillo, A. *et al.* Crosstalk between LXR and Toll-like receptor signaling mediates bacterial and viral antagonism of cholesterol metabolism. *Mol. Cell* **12**, 805–816 (2003).
- Lin, Y. *et al.* The lipopolysaccharide-activated Toll-like receptor (TLR)-4 induces synthesis of the closely related receptor TLR-2 in adipocytes. *J. Biol. Chem.* **275**, 24255–24263 (2000).
- Bookout, A. L. *et al.* Anatomical profiling of nuclear receptor expression reveals a hierarchical transcriptional network. *Cell* **126**, 789–799 (2006).
- Ogawa, S. *et al.* Molecular determinants of crosstalk between nuclear receptors and Toll-like receptors. *Cell* **122**, 707–721 (2005).
- Makowski, L. *et al.* Lack of macrophage fatty acid-binding protein aP2 protects mice deficient in apolipoprotein E against atherosclerosis. *Nature Med.* **7**, 699–705 (2001).
- Rosen, E. D. & Spiegelman, B. M. PPAR γ : a nuclear regulator of metabolism, differentiation, and cell growth. *J. Biol. Chem.* **276**, 37731–37734 (2001).
- Hotamisligil, G. S. *et al.* Uncoupling of obesity from insulin resistance through a targeted mutation in aP2, the adipocyte fatty acid binding protein. *Science* **274**, 1377–1379 (1996).
- Maeda, K. *et al.* Adipocyte/macrophage fatty acid binding proteins control integrated metabolic responses in obesity and diabetes. *Cell Metab.* **1**, 107–119 (2005).
- Tuncman, G. *et al.* A genetic variant at the fatty acid-binding protein aP2 locus reduces the risk for hypertriglyceridemia, type 2 diabetes, and cardiovascular disease. *Proc. Natl. Acad. Sci. USA* **103**, 6970–6975 (2006).
- Shum, B. O. *et al.* The adipocyte fatty acid-binding protein aP2 is required in allergic airway inflammation. *J. Clin. Invest.* **116**, 2183–2192 (2006).
- Makowski, L. & Hotamisligil, G. S. Fatty acid binding proteins — the evolutionary crossroads of inflammation and metabolic responses. *J. Nutr.* **134**, 2464S–2468S (2004).
- Taniguchi, C. M., Emanuelli, B. & Kahn, C. R. Critical nodes in signalling pathways: insights into insulin action. *Nature Rev. Mol. Cell Biol.* **7**, 85–96 (2006).
- White, M. F. IRS proteins and the common path to diabetes. *Am. J. Physiol. Endocrinol. Metab.* **283**, E413–E422 (2002).
- Hotamisligil, G. S. *et al.* IRS-1-mediated inhibition of insulin receptor tyrosine kinase activity in TNF- α - and obesity-induced insulin resistance. *Science* **271**, 665–668 (1996).
- Aguirre, V., Uchida, T., Yenush, L., Davis, R. & White, M. F. The c-Jun NH $_2$ -terminal kinase promotes insulin resistance during association with insulin receptor substrate-1 and phosphorylation of Ser³⁰⁷. *J. Biol. Chem.* **275**, 9047–9054 (2000).
- Paz, K. *et al.* Elevated serine/threonine phosphorylation of IRS-1 and IRS-2 inhibits their binding to the juxtamembrane region of the insulin receptor and impairs their ability to undergo insulin-induced tyrosine phosphorylation. *J. Biol. Chem.* **272**, 29911–29918 (1997).
- Emanuelli, B. *et al.* SOCS-3 inhibits insulin signaling and is up-regulated in response to tumor necrosis factor- α in the adipose tissue of obese mice. *J. Biol. Chem.* **276**, 47944–47949 (2001).
- Rui, L., Yuan, M., Frantz, D., Shoelson, S. & White, M. F. SOCS-1 and SOCS-3 block insulin signaling by ubiquitin-mediated degradation of IRS1 and IRS2. *J. Biol. Chem.* **277**, 42394–42398 (2002).
- Howard, J. K. *et al.* Enhanced leptin sensitivity and attenuation of diet-induced obesity in mice with haploinsufficiency of Socs3. *Nature Med.* **10**, 734–738 (2004).
- Gao, Z. *et al.* Serine phosphorylation of insulin receptor substrate 1 by inhibitor kappa B kinase complex. *J. Biol. Chem.* **277**, 48115–48121 (2002).
- Griffin, M. E. *et al.* Free fatty acid-induced insulin resistance is associated with activation of protein kinase C θ and alterations in the insulin signaling cascade. *Diabetes* **48**, 1270–1274 (1999).
- Hotamisligil, G. S. *et al.* IRS-1-mediated inhibition of insulin receptor tyrosine kinase activity in TNF- α - and obesity-induced insulin resistance. *Science* **271**, 665–668 (1996).
- Zick, Y. Ser/Thr phosphorylation of IRS proteins: a molecular basis for insulin resistance. *Sci. STKE* **2005** pe4 (2005).
- Baud, V. & Karin, M. Signal transduction by tumor necrosis factor and its relatives. *Trends Cell Biol.* **11**, 372–377 (2001).
- Hirosumi, J. *et al.* A central role for JNK in obesity and insulin resistance. *Nature* **420**, 333–336 (2002).
- Prada, P. O. *et al.* Western diet modulates insulin signaling, c-Jun N-terminal kinase activity, and insulin receptor substrate-1^{ser307} phosphorylation in a tissue-specific fashion. *Endocrinology* **146**, 1576–1587 (2005).

65. Ozcan, U. *et al.* Endoplasmic reticulum stress links obesity, insulin action, and type 2 diabetes. *Science* **306**, 457–461 (2004).
66. Tuncman, G. *et al.* Functional *in vivo* interactions between JNK1 and JNK2 isoforms in obesity and insulin resistance. *Proc. Natl Acad. Sci. USA* **103**, 10741–10746 (2006).
67. Ricci, R. *et al.* Requirement of JNK2 for scavenger receptor A-mediated foam cell formation in atherogenesis. *Science* **306**, 1558–1561 (2004).
68. Kaneto, H. N. Y. *et al.* Possible novel therapy for diabetes with cell-permeable JNK-inhibitory peptide. *Nature Med.* **10**, 1128–1132 (2004).
69. Liu, G. & Rondinone, C. M. JNK: bridging the insulin signaling and inflammatory pathway. *Curr. Opin. Investig. Drugs* **6**, 979–987 (2005).
70. Yuan, M. *et al.* Reversal of obesity- and diet-induced insulin resistance with salicylates or targeted disruption of I κ B. *Science* **293**, 1673–1677 (2001).
71. Hundal, R. S. *et al.* Mechanism by which high-dose aspirin improves glucose metabolism in type 2 diabetes. *J. Clin. Invest.* **109**, 1321–1326 (2002).
72. Cai, D. *et al.* Local and systemic insulin resistance resulting from hepatic activation of IKK- β and NF- κ B. *Nature Med.* **11**, 183–190 (2005).
73. Kim, J. K. *et al.* PKC- θ knockout mice are protected from fat-induced insulin resistance. *J. Clin. Invest.* **114**, 823–827 (2004).
74. Boden, G. *et al.* Free fatty acids produce insulin resistance and activate the proinflammatory nuclear factor-kappaB pathway in rat liver. *Diabetes* **54**, 3458–3465 (2005).
75. Liu, J. *et al.* NF- κ B is required for UV-induced JNK activation via induction of PKC δ . *Mol. Cell* **21**, 467–480 (2006).
76. Marciniak, S. J. & Ron, D. Endoplasmic reticulum stress signaling in disease. *Physiol. Rev.* **86**, 1133–1149 (2006).
77. Schroder, M. & Kaufman, R. J. The mammalian unfolded protein response. *Annu. Rev. Biochem.* **74**, 739–789 (2005).
78. Harding, H. P., Zhang, Y., Bertolotti, A., Zeng, H. & Ron, D. Perk is essential for translational regulation and cell survival during the unfolded protein response. *Mol. Cell* **5**, 897–904 (2000).
79. Harding, H. P., Zhang, Y. & Ron, D. Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase. *Nature* **397**, 271–274 (1999).
80. Urano, F. *et al.* Coupling of stress in the ER to activation of JNK protein kinases by transmembrane protein kinase IRE1. *Science* **287**, 664–666 (2000).
81. Deng, J. L. P. *et al.* Translational repression mediates activation of nuclear factor κ B by phosphorylated translation initiation factor 2. *Mol. Cell. Biol.* **24**, 10161–10168 (2004).
82. Hu, P. H. Z., Couvillon, A. D., Kaufman, R. J. & Exton, J. H. Autocrine tumor necrosis factor α links endoplasmic reticulum stress to the membrane death receptor pathway through IRE1 α -mediated NF- κ B activation and down-regulation of TRAF2 expression. *Mol. Cell. Biol.* **26**, 3071–3084 (2006).
83. Ozawa, K. *et al.* The endoplasmic reticulum chaperone improves insulin resistance in type 2 diabetes. *Diabetes* **54**, 657–663 (2005).
84. Nakatani, Y. *et al.* Involvement of endoplasmic reticulum stress in insulin resistance and diabetes. *J. Biol. Chem.* **280**, 847–851 (2005).
85. Ozcan, U. *et al.* Chemical chaperones reduce ER stress and restore glucose homeostasis in a mouse model of type 2 diabetes. *Science* **313**, 1137–1140 (2006).
86. Zhang, K. S. X. *et al.* Endoplasmic reticulum stress activates cleavage of CREBH to induce a systemic inflammatory response. *Cell* **124**, 587–599 (2006).
87. Cullinan, S. B. & Diehl, J. A. Coordination of ER and oxidative stress signaling: the PERK/Nrf2 signaling pathway. *Int. J. Biochem. Cell Biol.* **38**, 317–332 (2006).
88. Xue, X. P. J. *et al.* Tumor necrosis factor α (TNF α) induces the unfolded protein response (UPR) in a reactive oxygen species (ROS)-dependent fashion, and the UPR counteracts ROS accumulation by TNF α . *J. Biol. Chem.* **280**, 33917–33925 (2005).
89. Furukawa, S. *et al.* Increased oxidative stress in obesity and its impact on metabolic syndrome. *J. Clin. Invest.* **114**, 1752–1761 (2004).
90. Lin, Y. *et al.* The hyperglycemia-induced inflammatory response in adipocytes: the role of reactive oxygen species. *J. Biol. Chem.* **280**, 4617–4626 (2005).
91. Houstis, N., Rosen, E. D. & Lander, E. S. Reactive oxygen species have a causal role in multiple forms of insulin resistance. *Nature* **440**, 944–948 (2006).
92. Brownlee, M. Biochemistry and molecular cell biology of diabetic complications. *Nature* **414**, 813–820 (2001).
93. Lowell, B. B. & Shulman, G. I. Mitochondrial dysfunction and type 2 diabetes. *Science* **307**, 384–387 (2005).
94. Steen, E. *et al.* Impaired insulin and insulin-like growth factor expression and signaling mechanisms in Alzheimer's disease — is this type 3 diabetes? *J. Alzheimers Dis.* **7**, 63–80 (2005).
95. Mootha, V. K. *et al.* PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet.* **34**, 267–273 (2003).
96. St-Pierre, J. *et al.* Suppression of reactive oxygen species and neurodegeneration by the PGC-1 transcriptional coactivators. *Cell* **127**, 397–408 (2006).
97. Gao, Z., Zuberi, A., Quon, M. J., Dong, Z. & Ye, J. Aspirin inhibits serine phosphorylation of insulin receptor substrate 1 in tumor necrosis factor-treated cells through targeting multiple serine kinases. *J. Biol. Chem.* **278**, 24944–24950 (2003).
98. Waeber, G. *et al.* The gene MAPK8IP1, encoding islet-brain-1, is a candidate for type 2 diabetes. *Nature Genet.* **24**, 291–295 (2000).
99. Klier, S. A., Xu, H. E., Lambert, M. H. & Wilson, T. M. Peroxisome proliferator-activated receptors: from genes to physiology. *Recent Prog. Horm. Res.* **56**, 239–263 (2001).
100. Mahata, B., Mukherjee, S., Mishra, S., Bandyopadhyay, A. & Adhya, S. Functional delivery of a cytosolic tRNA into mutant mitochondria of human cells. *Science* **314**, 471–474 (2006).

Acknowledgements I am grateful for the contributions of many students and fellows for studies performed in my laboratory. My work is supported by the National Institutes of Health, American Diabetes Foundation, Pew Foundation, Sandler Foundation, Iacocca Foundation and Harvard School of Public Health. I particularly thank K. Freidinger Wellen, M. Gustafson Gregor, U. Ozcan, G. Tuncman, M. Furuhashi and R. Foote for their contributions to this review. Special thanks to D. Hotamisligil for help with illustrations. I apologize to my colleagues whose work I was unable to cite because of space limitations.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to the author (ghotamis@hsph.harvard.edu).

Sirtuins as potential targets for metabolic syndrome

Leonard Guarente¹

Metabolic syndrome threatens health gains made during the past century. Physiological processes degraded by this syndrome are often oppositely affected by calorie restriction, which extends lifespan and prevents disease in rodents. Recent research in the field of ageing has begun to identify important mediators of calorie restriction, offering the hope of new drugs to improve healthspan. Moreover, if metabolic syndrome and calorie restriction are opposite extremes of the same metabolic spectrum, calorie restriction mimetics might provide another therapeutic approach to metabolic syndrome. Sirtuins and other important metabolic pathways that affect calorie restriction may serve as entry points for drugs to treat metabolic syndrome.

In the western world the prevalence of metabolic syndrome in the adult population is approaching one-quarter, probably triggered by high-calorie diets and physical inactivity. It is characterized by a combination of physiological parameters, including obesity, inflammation, high blood pressure and dyslipidaemia (high levels of circulating triacylglycerols and low-density lipoprotein (LDL) cholesterol, and low levels of high-density lipoprotein (HDL) cholesterol)^{1–3}. Metabolic syndrome is also associated with dysregulation of glucose homeostasis — that is, glucose intolerance (the inability to clear an orally administered dose of glucose from the blood normally), which is indicative of insulin insensitivity (inability of insulin to promote normal glucose uptake by cells). This dysregulation can be associated with higher levels of blood insulin — a compensation mechanism — and, as the syndrome progresses, increased blood glucose levels and diabetes. Metabolic syndrome was first recognized as a risk factor for cardiovascular disease, and is associated with atherosclerosis. This syndrome also heightens risk for stroke, cancer, arthritis and, of course, diabetes. Lifestyle changes are the first defence in treating metabolic syndrome, followed by pharmacological intervention.

Whereas prediabetic conditions were once thought to be related to ageing, as are type II diabetes and cardiovascular disease, the recent epidemic of metabolic syndrome has afflicted younger adults and even children. Nevertheless, there does seem to be an ageing- or time-dependent component to the progression from metabolic syndrome to diabetes, and the resulting high risk for cardiovascular disease. Moreover, a link can be imagined between metabolic syndrome and our evolutionary strategy for survival.

It is likely that the selected evolutionary strategy in times of food availability was the preferential use of carbohydrates for energy, and the storage of fat, because fat is more reduced and has a higher energy content per unit mass. Thus, animals may have taken advantage of the fact that fat storage was a sign that food was available and leanness was a sign of food scarcity. More specifically, fat cells are known to secrete hormones known as adipokines, so this dietary information could readily be disseminated throughout the body. In times of food availability, the best life strategy would be to reproduce and not worry about future, post-reproductive health deterioration. In times of food scarcity, the opposite strategy would apply. In the western world, where food is abundant, we may therefore be harvesting the consequences of an evolutionary strategy that neglected the long-term health effects of caloric excess.

In this review I explore how recent findings in the study of ageing might have implications for understanding and treating metabolic syndrome. In particular, I focus on the link between SIR2-related proteins (sirtuins) and calorie restriction (CR), and present a hypothesis that metabolic syndrome and CR might lie at opposite ends of the same spectrum. Therefore, findings on how CR works may provide new possibilities for treating metabolic syndrome.

Calorie restriction and metabolic syndrome

Calorie restriction was first described as a reduction in food intake in laboratory rodents of between 20% and 40% of *ad libitum* levels that would extend their lifespan by up to 50%⁴. It now seems that CR works universally to promote survival in organisms ranging from yeast to rodents and, perhaps, primates. As described above, CR may have evolved as an adaptive trait to postpone reproduction during food scarcity to a later time of food availability⁵. If CR thus evolved as a programme, it may be regulated by a relatively small number of genes. Recent findings have linked CR to the *SIR2* gene family, which were first shown to have anti-ageing functions in yeast⁶, *Caenorhabditis elegans*⁷ and *Drosophila*⁸. The discovery that yeast Sir2 and the mammalian orthologue SIRT1 are NAD⁺-dependent deacetylases^{9,10} spurred the hypothesis that sirtuins might regulate the pace of ageing in accord with metabolism, and might therefore provide the longevity that results from CR.

Do CR and metabolic syndrome lie at opposite ends of the same spectrum and so involve an overlapping set of regulators? Several considerations suggest that this may be the case. First is the obvious fact that metabolic syndrome is triggered by dietary excess and CR by dietary restriction. Second, many of the physiological parameters that are characteristic of metabolic syndrome (described above) are oppositely affected by CR, which yields improved glucose tolerance (and lower blood glucose and insulin levels), decreased LDL cholesterol and triacylglycerols, and increased HDL cholesterol. Third, whereas metabolic syndrome predisposes to diseases, CR protects against many diseases in rodent models, including cardiovascular disease, cancer, diabetes and neurodegenerative disease^{11–13}.

Thus, it may be useful to think of metabolic syndrome and CR as lying at opposite ends of a balance, which can be tipped in either direction by diet and physical activity (Fig. 1). Most importantly, this hypothesis

¹Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

posits that the regulatory factors that mediate the positive effects of a low-calorie diet may also have direct relevance to at least the glucose intolerance and obesity of metabolic syndrome. Below, I focus on a group of such factors — the sirtuins — and also discuss the transcriptional coactivators PPAR- γ (peroxisome-proliferator-activated receptor- γ) coactivator-1 α (PGC-1 α) and PGC-1 β that are involved in regulating metabolic genes in the liver, muscles and brown fat, and AMP-activated protein kinase (AMPK), which is normally activated in many cell types by a deficit in energy.

Calorie restriction in various organisms

Studies on ageing in yeast mother cells show that Sir2 has at least two activities that might promote longevity for mothers and also confer fitness on daughter cells. First, it represses genome instability in the rDNA repeats and thus slows the formation of toxic rDNA circles¹⁴. Second, it promotes the asymmetric segregation of oxidatively damaged proteins to mother cells, and thereby resets the full lifespan to the damage-free daughters¹⁵.

A regimen for CR in yeast was described in which mother cells were grown on 0.5% glucose as their carbon and energy source, instead of the usual 2% glucose¹⁶. Under these conditions of moderate CR, knocking out *SIR2* alone prevented lifespan extension in some yeast strains¹⁷, whereas knocking out *SIR2* and two *SIR2* paralogues was required to block the extension in another strain¹⁸. Importantly, these effects were observed in strains that also bore deletions in *FOB1*, which prevented the accumulation of rDNA circles and their accompanying short lifespan in Sir2 mutants¹⁷.

Two mechanisms have been shown to upregulate Sir2 activity during the moderate 0.5% glucose CR regimen (Fig. 2). In the first, CR was shown to trigger a metabolic shift from fermentation to respiration, and this increase in respiration was required for life extension¹⁷. Higher respiration rates resulted in an increase in the NAD⁺/NADH ratio and the corresponding activation of Sir2 (ref. 19). In the second, CR was shown to upregulate *PCN1* — which re-synthesizes NAD⁺ from nicotinamide and ADP-ribose — and thereby lower the levels of nicotinamide, a potent Sir2 inhibitor²⁰. A more severe 0.05% glucose CR regimen also extended the lifespan of mother cells, but in a manner not requiring Sir2 and perhaps invoking the TOR nutrient-sensing pathway^{21,22}.

In *Drosophila*, CR — achieved by means of a modest reduction in the yeast extract in food — was shown to reduce the expression of the general histone deacetylase, RPD3, which, in turn, resulted in an increase in *SIR2* mRNA expression²³. Moreover, knocking out *Sir2* prevented the longevity induced by CR, and both *Sir2* overexpression and CR gave lifespan extensions that were not additive^{8,24}. Finally, in *C. elegans*, the extension in lifespan in *eat* mutants, which are defective in pharyngyl pumping of food, seemed to be at least partly dependent on *sir-2.1* (ref. 25). These findings all suggest a key role for sirtuins in mediating effects of moderate CR in lower organisms.

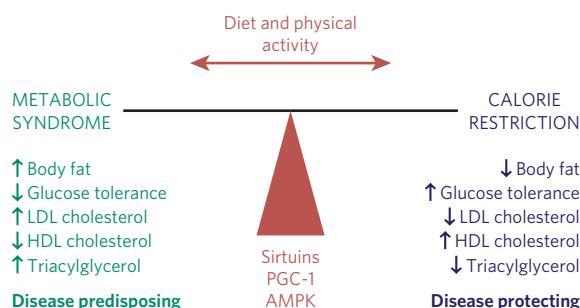


Figure 1 | Metabolic syndrome and calorie restriction are balanced at opposite ends of the same spectrum by diet and physical activity. The regulators shown might be involved in the underlying mechanisms that influence the balance. The reciprocity of phenotypes of metabolic syndrome and calorie restriction and their effects on disease are also indicated.

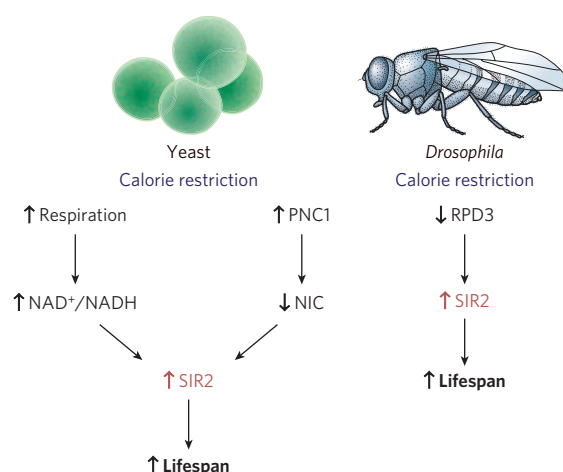


Figure 2 | Pathways of SIR2 activation by moderate calorie restriction in yeast and *Drosophila*. In yeast, two pathways activate SIR2 during CR, one involving an increase in respiration and the NAD⁺/NADH ratio, the other an increase in the NAD⁺-scavenging pathway enzyme, PCN1, which reduces nicotinamide (NIC) levels. In *Drosophila*, CR represses expression of the class I deacetylase RPD3, thereby activating *Drosophila* SIR2.

Are mammalian sirtuins required for CR-induced effects? In at least one example, the answer seems to be yes. CR mice showed a large increase in physical activity that seems to require SIRT1, because the increase did not occur in *Sirt1*-knockout mice²⁶. Also, many of the functions described below for SIRT1, 3, 4 and 7 are consistent with a role for mammalian sirtuins in CR-induced changes in metabolism and increases in stress tolerance.

Functions of SIRT1 in mammalian physiology

The initial characterization of SIRT1 showed that it deacetylates important transcription factors, including p53, forkhead subgroup O (FOXO) proteins and the DNA repair factor KU, thereby increasing the stress resistance of cells by inhibiting apoptosis and increasing repair^{27–32}. Moreover, SIRT1 has been linked to both lipid and glucose homeostasis. In white adipose tissue, SIRT1 was shown to inhibit adipogenesis in precursor cells and to reduce fat storage in differentiated cells³³. One mechanism involved seemed to be inhibition of the nuclear receptor, PPAR- γ , by SIRT1 docking with its negative cofactors NCOR and SMRT at target gene promoters. However, because this mechanism does not explain the lipolysis triggered by CR in adipocytes, other activities may also be important.

SIRT1 can also regulate glucose homeostasis in three different tissues by affecting different targets (Fig. 3). In pancreatic β -cells, SIRT1 is a positive regulator of insulin secretion^{34,35}. Insulinoma cells with SIRT1 reduced by RNA inhibition showed impaired insulin secretion, and transgenic mice overexpressing SIRT1 specifically in β -cells had improved glucose tolerance. Lowering SIRT1 in the insulinoma cells activated transcription of the uncoupling protein 2 gene (*Ucp-2*), whereas the SIRT1 transgenic mice showed super-repressed levels of UCP-2. Because UCP-2 encodes a mitochondrial membrane protein that might uncouple ATP synthesis from respiration, its repression by SIRT1 may increase the efficiency of ATP synthesis in β -cells in response to glucose, and thus positively regulate insulin secretion.

SIRT1 was also shown to protect β -cells against oxidative stress in a mechanism proposed to involve deacetylation of FOXO proteins³⁶. So this sirtuin might also restrain β -cell loss during ageing and thereby mitigate a catastrophic reduction in insulin production in patients with early-stage diabetes to slow the progression to full-blown disease.

In the liver, SIRT1 seems to regulate gluconeogenesis. In liver cells, this sirtuin bound to and deacetylated the PPAR- γ coactivator PGC-1 α ³⁷ (discussed in detail below), thereby activating it. Indeed, SIRT1 levels in the liver were shown to increase markedly after overnight fasting,

resulting in an increase in glucose production. Because PGC-1 α and FOXO proteins both regulate genes involved in gluconeogenesis, there are clearly several mechanisms by which SIRT1 could affect glucose production in the liver in times of severe energy limitation. In neurons, SIRT1 seemed not to activate but to repress the activity of PGC-1 α ³⁸, revealing the complexity of PGC-1 α regulation by this sirtuin.

Finally, SIRT1 might also affect glucose homeostasis by regulating the response of target cells (such as muscle cells) to insulin. This hormone activates a pathway of intracellular kinases that regulate forkhead transcription factors^{39,40}, which, as mentioned above, are directly regulated by SIRT1. Moreover, PGC-1 α activates genes involved not only in gluconeogenesis, but also in mitochondrial biogenesis, fatty acid oxidation and respiration (see below). By regulating the activity of PGC-1 α in the muscles and liver, SIRT1 may also influence the abilities of these tissues to respire and metabolize carbohydrates and fats. The regulation of PGC-1 α by SIRT1 could thus influence both glucose and lipid homeostasis.

SIRT1 activity during calorie restriction

Does SIRT1 activity increase in all tissues during food limitation? The first indication that the answer to this question may well be no was the finding that fasting in wild-type but not *Sirt1*-knockout mice increased pancreatic UCP-2, implying that a reduction in SIRT1 activity occurred during fasting³⁴. Consistent with this was the finding that the NAD⁺/NADH ratio decreased in starved pancreas, whereas the NAD⁺/NADH ratio in the liver increased after fasting³⁷. Thus, during periods of acute food shortage, it seems possible that the activity of SIRT1 changes in different directions in different tissues. However, one caveat is that the NAD⁺/NADH ratio has not clearly been shown to be the primary determinant of SIRT1 activity in mammals, as opposed to, for example, changes in protein levels.

SIRT1 can increase the stress resistance of cells, so during long-term CR its activity might be expected to rise in all tissues. Indeed, SIRT1 protein levels have been shown to increase during CR in the brain, white adipose tissue, muscles, liver and kidneys^{33,41}. However, a decrease in the NAD⁺/NADH ratio in the livers of mice during CR has also been reported⁴². A functional assay, described below, was consistent with this latter finding, showing that the activity of another sirtuin, SIRT4, also decreases in the liver during CR. Given these disparate observations, it will be important to study the liver in more detail — for example, comparing transcription profiles of wild-type and *Sirt1*-knockout mice — to determine whether SIRT1 activity rises or falls during CR. If it turns out that SIRT1 activity does change in different directions in different tissues during CR, as seems to be the case in fasting, then pharmacological

interventions that activate or repress sirtuins in the whole animal may mimic CR in only a segmental fashion.

Most importantly, in line with the model in Fig. 1, it will be important to determine whether the changes in SIRT1 activity during CR are inverse to changes observed in genetically or dietary-induced obese animals. If so, SIRT1 and perhaps other sirtuins could be potential pharmacological targets not only for diseases of ageing but also for metabolic syndrome.

Mitochondrial SIRT3 and SIRT4 in metabolism

Mitochondria have figured prominently in at least some models of ageing⁴³, such as the oxidative damage theory, which proposes that reactive oxygen species generated as a by-product of respiration cause cumulative damage in mitochondria. In addition to providing the 'factory' for respiration and ATP production, mitochondria also house many metabolic pathways. The fact that both SIRT3 and 4 are imported into the mitochondrial matrix^{44–46} suggests that these sirtuins might have a role in stress management and metabolism. Although the role of mitochondria in ageing remains putative, the fact that SIRT3, 4 and 5 have all been reported to be mitochondrial proteins provides further support for the potential importance of this organelle in ageing.

SIRT3 was recently shown to deacetylate the mitochondrial enzyme acetyl-coenzyme-A synthetase 2 (AceCS2)^{47,48}, which converts acetate to acetyl-CoA, thereby allowing the entry of carbon from dietary acetate into central metabolism (Fig. 4). This is a strikingly conserved function, because the sole bacterial sirtuin, CobB, was shown to deacetylate bacterial AceCS¹⁹. Because the acetylated lysine in AceCS is in the active site, deacetylation activates the enzyme. Notably, CobB is required for bacteria to use acetate as a carbon source. In mammals, whereas SIRT3 deacetylated and activated AceCS2, SIRT1 was reported to deacetylate and activate the cytoplasmic isoform, AceCS1 (ref. 47). Although many studies have shown that SIRT1 is nuclear, its presence in the cytoplasm has been reported in some cell types under certain conditions³⁵. These findings all suggest that SIRT3 (and perhaps SIRT1) might regulate the entry of acetate into the tricarboxylic acid cycle and central metabolism. This step might be especially important during times of food limitation in order to both harvest dietary acetate and make use of the acetate that is known to be generated by the liver during ketogenesis⁵⁰. It will be important to demonstrate directly the physiological relevance of these biochemical findings — for example, by studying the effects of different diets in *Sirt3*^{-/-} mice.

SIRT4 also regulates the flow of carbon into central metabolism, in this case from the amino acids glutamate and glutamine (Fig. 4). Biochemical studies of SIRT4 showed that it does not have NAD⁺-dependent deacetylase activity, but instead uses NAD⁺ to transfer ADP-ribose to protein substrates⁴⁶. The physiologically relevant substrate for this ADP-ribosyltransferase activity turned out to be the mitochondrial enzyme glutamate dehydrogenase (GDH). By ADP-ribosylating GDH, SIRT4 inhibits its activity and blocks the conversion of glutamate (and glutamine, which is converted to glutamate in cells) to the tricarboxylic acid cycle intermediate, α -ketoglutarate.

Importantly, pancreatic β -cells were found to be highly enriched in SIRT4, and knocking out *Sirt4* in both insulinoma cells and mice triggered insulin hypersecretion⁴⁶. This increase seems to be due to the potential use of these amino acids as fuel sources in β -cells lacking SIRT4. Indeed, unlike the wild type, the *Sirt4*-knockout mice secreted insulin in response to glutamine as well as glucose. Thus, SIRT4 functions to repress amino-acid-stimulated insulin secretion (AASIS) in β -cells.

The physiological role of SIRT4 becomes clear when it is considered that amino acids can serve as carbon and energy sources in times of energy limitation. The β -cells of wild-type mice on a CR diet have been shown to secrete insulin in response to glutamine⁴⁶. This qualitative change in insulin responsiveness seemed to be due to downregulation of SIRT4, because GDH was less ADP-ribosylated in mitochondria from CR mice than in those of controls. Similarly, in the liver, GDH was less ADP-ribosylated in CR mice, which would allow for the use of amino

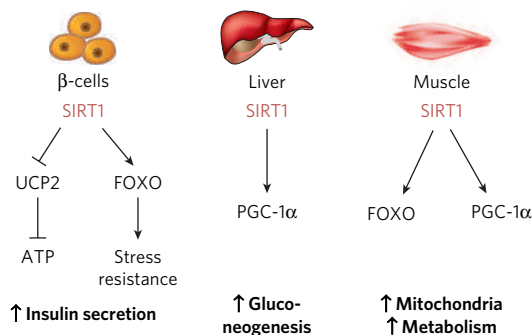


Figure 3 | Influence of SIRT1 on glucose homeostasis in three mammalian tissue types. In β -cells, SIRT1 represses the uncoupling protein gene, *Ucp2*, and thereby increases ATP synthesis and insulin secretion in response to glucose. SIRT1 also protects β -cells against stress-induced apoptosis by increasing activity of the forkhead protein FOXO1. In the liver, SIRT1 deacetylates the coactivator PGC-1 α , thereby increasing expression of genes for gluconeogenesis. In the muscles, the effect of SIRT1 on FOXO1 and PGC-1 α proteins should result in an increase in mitochondrial biogenesis and metabolism.

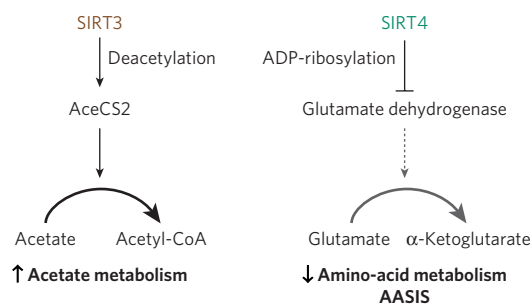


Figure 4 | Functions of SIRT3 and SIRT4 in regulating the entry of acetate or amino acids into central metabolism. SIRT3 deacetylates and activates the mitochondrial enzyme AceCS2, which converts acetate to acetyl-CoA, thereby facilitating use of acetate in metabolism. SIRT4 ADP-ribosylates the mitochondrial enzyme glutamate GDH, which converts glutamate to α -ketoglutarate, thereby repressing the entry of glutamate and glutamine into metabolism and blocking their ability to trigger AASIS.

acids for gluconeogenesis. Thus, whether glutamine and glutamate can be used as fuel sources in central metabolism and in AASIS is regulated by SIRT4 according to diet.

It is fascinating to note that SIRT3 and 4 seem to function oppositely with respect to carbon use — SIRT3 promotes the use of acetate, whereas SIRT4 represses the use of glutamate and glutamine. Because both SIRT3 and 4 are likely to be regulated in the same direction in the same cellular compartment by changes in the NAD^+/NADH ratio, their roles seem to be conflicting.

How can we make sense of this? I speculate that the ability to use one or other fuel source during CR is parsed between different tissues (Fig. 5). For example, the metabolism of amino acids to make glucose clearly occurs in the liver. Because some of the amino acids used for gluconeogenesis come from protein breakdown in the muscles, it would make sense to downregulate SIRT4 specifically in the liver to increase GDH activity and amino-acid metabolism (Fig. 5). As amino-acid metabolism generates glucose under these conditions, we can begin to understand teleologically the qualitative shift to AASIS in β -cells, which is also mediated by downregulation of SIRT4.

In a reciprocal fashion, it may be desirable to potentiate the use of dietary acetate as a carbon and energy source in the muscles but not the liver, where it is produced during ketogenesis. Consistent with this idea, AceCS2 is abundant in skeletal muscle and the heart, but almost absent from the liver, and is highly upregulated in muscles during food limitation⁵¹. Whether SIRT3, which is expressed at very low levels in the muscles, is highly induced in muscle tissue during ketogenic conditions remains to be tested. If so, SIRT3 and 4 might have reciprocal systemic roles in the liver and muscles during food limitation to facilitate the use by each tissue of a fuel source sent by the other for metabolism and generation of energy.

In summary, SIRT3 and 4 clearly have important roles in diet-induced metabolic changes. Because mitochondria are so important in stress and, perhaps, ageing as generators of energy, recipients of damage and regulators of apoptosis⁴³, it will be interesting to see whether the mitochondrial sirtuins also function in stress management.

Nuclear SIRT6 and SIRT7 and metabolism

Along with SIRT1, SIRT6 and 7 are the other nuclear sirtuins. Interesting functions for SIRT6 emerged from the analysis of the *Sirt6*-knockout mouse, which exhibits genomic instability and a progeroid phenotype⁵². A defect in base excision repair was found that might explain the cell loss leading to the rapid ageing phenotype. However, the mice also showed severe defects in glucose homeostasis and low levels of insulin-like growth factor (IGF-1), which were evident even before the onset of the degenerative phenotypes. In fact, the attrition in lymphocytes that was observed in these mice resulted from a systemic effect, perhaps the defect in glucose homeostasis or IGF-1. Thus, like

SIRT1, SIRT6 might have an important role in glucose homeostasis, and further studies should provide important information about whether this sirtuin helps coordinate metabolic changes with diet.

SIRT7 is the only sirtuin shown to be localized in nucleoli^{53,54}, where it is associated with RNA polymerase I (ref. 54). Indeed, SIRT7 seems to be a positive regulator of rRNA transcription, because its inhibition reduced transcription and its overexpression enhanced it. However, regulation of ribosome biogenesis by sirtuins may be more complex, as SIRT1 has been reported to deacetylate the RNA polymerase factor TAF168 and thereby regulate rRNA transcription in the opposite direction⁵⁵. SIRT7 is highly expressed in many tissues with dividing cells⁵⁴. It will be of interest to determine whether SIRT7 activity decreases in these tissues during CR to restrain ribosome biogenesis and cell growth when energy is limiting. By contrast, SIRT7 may not have an important role in organs consisting of postmitotic cells such as the muscles, heart and brain, because expression of this sirtuin was not observed in these tissues.

Possible links between sirtuins and metabolic syndrome

Because of the properties of SIRT1, 3 and 4 outlined above, it might be useful to consider possible effects on metabolic syndrome of activating or inhibiting these sirtuins in different tissues (Table 1). The cases of SIRT3, 4 and 1 seem to provide examples of increasing complexity. Both SIRT3 and 4 regulate the flow of carbon from acetate and amino acids into metabolism, through which they could contribute to the synthesis of carbohydrate or fat. It may be useful, therefore, to inhibit SIRT3 to block any incorporation of acetate into metabolism for synthesis. The same logic applies to SIRT4, except that in this case it is activation that would reduce entry of glutamate and glutamine into central metabolism — for example, as fuel for gluconeogenesis in the liver. However, this sirtuin may be more complex than SIRT3 — it is the inhibition of SIRT4 in β -cells that might provide at least temporary benefit for glucose intolerance, because it would increase AASIS.

In the case of SIRT1, it seems likely that activation in white adipose tissue would provide benefit by stimulating fat loss. Likewise, activation in β -cells might help early-stage diabetes by increasing insulin production (Table 1). We can also speculate on a beneficial role for SIRT1 activation in muscle to provide stress resistance and prevent muscle loss. However, we will not know whether activation or inhibition of SIRT1 in the liver is useful until we know whether the effects on this tissue observed during fasting apply to long-term CR. It should be possible to test whether regulating SIRT1, 3 and 4 in the indicated directions and tissues brings about the desired effects by generating tissue-specific knockout and transgenic mice for these genes. If such genetically altered mice demonstrate an improved physiological response when challenged with diets high in

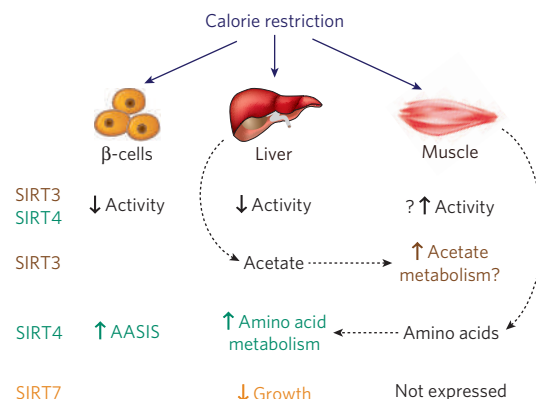


Figure 5 | Model of the effects of SIRT3, SIRT4 and SIRT7 in different tissues during calorie restriction. The indicated direction of change in sirtuin activity is the best surmised on the basis of published data. Question marks indicate that SIRT3 has not yet been tested for induction by CR in muscle. Note the reciprocal effects of SIRT3 and SIRT4 on metabolism of acetate and amino acids in muscle and liver.

fat and carbohydrate, it can be hypothesized that manipulating these sirtuins might benefit humans with metabolic syndrome. However, the pathway to developing drugs to selectively activate or repress a specific sirtuin in a particular tissue will be considerably more challenging than the genetic proof of principle studies.

PGC-1 α and PGC-1 β

PGC-1 α was identified as a coactivator that bound to the nuclear receptor, PPAR- γ , and stimulated fat metabolism and thermogenesis in brown fat cells⁵⁶. This protein has other important roles in the muscles and liver during energy limitation that might be relevant to metabolic syndrome and make PGC-1 proteins attractive targets⁵⁷. In muscles, exercise can induce the β -adrenergic system to activate cyclic-AMP (cAMP)-dependent protein kinase and its transcription factor target, cAMP-responsive element-binding protein (CREB) to upregulate PGC-1 α expression. This increase can then drive differentiation of slow twitch fibres, which, unlike fast twitch fibres, make exclusive use of oxidative metabolism for energy production. In these fibres, PGC-1 α stimulates transcription of nuclear genes encoding mitochondrial proteins by binding to transcription factors such as nuclear respiratory factors 1 and 2 (NRF-1/2) and oestrogen-related receptor (ERR) proteins. PGC-1 α also activates fatty acid oxidation by binding to PPAR- α and δ . The net effect of PGC-1 α activity in muscle is therefore an increase in fatty acid oxidation and metabolic activity. Furthermore, PGC-1 α mRNA has been shown to be decreased in the muscles of patients with type 2 diabetes^{58,59}, although it is not yet clear whether this change contributes to disease pathology. Thus, it is a reasonable deduction that activation of PGC-1 α in muscle could provide benefit for metabolic syndrome (Table 1).

In the liver, PGC-1 α was shown to activate both fatty acid oxidation and gluconeogenesis by binding to transcription factors FOXO1 and hepatocyte nuclear factor-4 α (HNF4 α)⁶⁰. Thus, logic might suggest that inhibiting its activity in this tissue might help slow the progression from glucose intolerance to diabetes in people with metabolic syndrome (Table 1). One possible complication, however, is that PGC-1 α inhibition could lead to steatosis, or fatty liver, due to compromised fat oxidation. This would reduce hepatic insulin sensitivity, thereby countering some of the beneficial effects on glucose output and perhaps leading to other hepatic problems.

In this same tissue, PGC-1 β was shown to activate cholesterol and fat synthesis and export to the bloodstream by binding to the lipogenic transcription factors sterol regulatory element binding protein (SREBP) and liver X receptor (LXR)⁶¹. Therefore, inhibiting PGC-1 β in the liver might be of benefit in ameliorating the hyperlipidaemia in patients with metabolic syndrome. However, another report shows that PGC-1 β is a coactivator for the forkhead protein FOXA2 (ref. 62). Forkhead proteins are normally repressed by insulin signalling, because it leads to their phosphorylation by AKT (also known as protein kinase B) and retention in the cytoplasm. Indeed, fasting was shown to promote the nuclear localization of hepatic FOXA2, where it increased fatty acid oxidation, glycolysis and ketogenesis, and reduced gluconeogenesis and hepatic fat⁶³. These properties suggest that it is the activation of PGC-1 β that would cause a hepatic response favourable for metabolic syndrome. Further study will be required to resolve which set of these apparently opposing activities of PGC-1 β is most relevant to metabolic syndrome.

So, both SIRT1 and PGC-1 proteins probably have important roles in muscles and the liver (Table 1). The function of sirtuins may be broader and encompass white adipose tissue, β -cells and probably other tissues as well. Both SIRT1 and PGC-1 α are upregulated by energy limitation⁴¹, and thereby exert coordinated effects in the liver and muscles during a state of food limitation — for example, upregulation of fatty acid oxidation to provide carbon for gluconeogenesis. However, in the face of energy excess, it might be most efficacious to activate oxidative metabolism in order to reduce fat, but to avoid activating gluconeogenesis, which would exacerbate a pre-diabetic condition. Achieving this aim by pharmacologically modulating PGC-1 proteins or the transcription

Table 1 | Functions of various regulators in β -cells, the liver and muscles

Regulator		β -cell	Liver	Muscle
SIRT1	Activity	Increases insulin secretion	Metabolism	Increases stress resistance
	Therapy	Activate	None known*	Activate
SIRT3	Activity			Increases acetate metabolism
	Therapy			Inhibit
SIRT4	Activity	Decreases AASIS	Decreases amino-acid metabolism	
	Therapy	Inhibit	Activate	
PGC-1 α	Activity		Increases gluconeogenesis	Increases metabolism
	Therapy		Inhibit†	Activate
PGC-1 β	Activity		Increases fat/cholesterol	
	Therapy		Inhibit†	
AMPK	Activity		Decreases fat/cholesterol	Increases fat oxidation
	Therapy		Activate	Activate

*The role of SIRT1 in the liver in CR is not fully understood.

†May involve complications.

Therapy for metabolic syndrome is predicted as activation or inhibition of the indicated sirtuin in the indicated tissue.

factors through which they function stands as an important challenge in devising new treatments for insulin insensitivity and obesity.

AMP-activated protein kinase

Another intriguing regulator of energy homeostasis is AMP-activated protein kinase (AMPK), which senses the AMP/ATP ratio in cells^{64,65}. During energy or food limitation, AMP binds to AMPK and renders it a substrate for the activating kinase LKB1 (refs 66–68). In neurons, another Ca²⁺/calmodulin-sensitive kinase also phosphorylates AMPK on the same residue without the requirement for bound AMP⁶⁹.

AMPK is already a prime target for treatment of metabolic syndrome, because one leading drug currently in use, metformin, is thought to work by activating this kinase, although the mechanism is not certain⁷⁰. Although it is beyond the scope of this review to cover all of the known effects of AMPK, several of its targets seem especially pertinent. First, AMPK phosphorylates and inhibits acetyl-CoA carboxylase, which converts acetyl-CoA to malonyl-CoA⁷¹. The product of this reaction is the building block for fatty acid synthesis in the liver. Malonyl-CoA also blocks fatty acid oxidation in muscles by inhibiting its transport into mitochondria^{64,65}. So, activating AMPK leads to inhibition of fatty acid synthesis in the liver and promotion of fatty acid oxidation in muscles (Table 1). Second, AMPK phosphorylates and inhibits 3-hydroxy-3-methylglutaryl-CoA reductase^{64,65}, which catalyses the committed step in cholesterol synthesis in the liver, so activation of AMPK also leads to a decrease in cholesterol production. Third, AMPK activates the PGC-1 α promoter, and its activation will thereby increase metabolism in muscles, as discussed above.

Finally, recent studies have identified another pathway that is relevant to both AMPK and PGC-1 α activity in the liver. TORC2 (CREB-regulated transcription coactivator 2) is induced by fasting to enter the nucleus and coactivate CREB, along with the canonical CREB coactivator CBP⁷². TORC2 seems to be especially important in triggering the activation of gluconeogenesis, probably by helping CREB to upregulate expression of PGC-1 α , as described above. In addition, nuclear TORC2 triggers a feedback mechanism in which it upregulates expression of the insulin pathway protein, insulin receptor substrate 2 (IRS2), to improve insulin signalling and temper gluconeogenesis⁷³. Most importantly, TORC2 can be phosphorylated by AMPK or the related kinase SIK to return it to its inactive, cytoplasmic state⁷². Indeed, knocking out the serine/threonine kinase LKB1 (and thus AMPK activity) in the liver activated TORC2, thereby driving PGC-1 α expression and gluconeogen-

esis⁷⁴. Thus, the activation of AMPK may also reduce gluconeogenesis in the liver by inhibiting TORC2.

Although at present there are no known direct connections between AMPK and sirtuins, it would not be surprising to see their emergence, given their common use in adapting an animal's metabolism to the energy needs imposed by its diet. One possible and intriguing intersection would be the regulation of one or more of the AMPK kinases by a sirtuin.

Summary and conclusion

Metabolic syndrome is a major health challenge of the twenty-first century, threatening to reverse historic trends towards ever increasing life- and healthspans in the developed world. We are on the cusp of a molecular understanding of ageing itself, and how it is regulated by diet. This research has dovetailed with studies of obesity, diabetes and metabolic disease to introduce us to some of the critical regulators of metabolic functions in mammals. In this review I have focused on the sirtuins, because they are candidates for regulators that bridge the control of metabolism and ageing. Because of this they, along with other metabolic regulators such as PGC-1 proteins, AMPK, FOXA2 and TORC2, are likely to be important to our understanding of how a low-calorie diet — that is, calorie restriction — promotes longevity and disease resistance. But equally importantly, they might provide insight into metabolic syndrome, because these same regulators may go awry in this pathological state. For this reason, it is possible that the development of drugs that target these metabolic regulators will not only be useful in combating ageing and its associated diseases but will also be effective in treating the insulin insensitivity, obesity and perhaps other symptoms associated with metabolic syndrome. Thus, we can imagine new treatments for diabetes, cardiovascular disease and other ageing-associated diseases that begin well before the onset of any noticeable symptoms. Like low-dose aspirin and the statins, such a class of drugs may vastly improve quality of life and productivity in an ageing cohort of people. Although some have questioned the ethics of anti-ageing research, its potential to mitigate metabolic syndrome and diseases of ageing demands that it proceed as rapidly as possible. ■

Note added in proof: Two recent studies have shown that the plant polyphenol resveratrol activates SIRT1 and mitigates effects of high-calorie and high-fat diets in mice^{75,76}.

- Luchsinger, J. A. A work in progress: the metabolic syndrome. *Sci. Aging Knowl. Environ.* **10**, pe19 (2006).
- Grundey, S. M. *et al.* Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute scientific statement. *Circulation* **112**, 2735–2752 (2005).
- Wilson, P. W. F., D'Agostino, R. B., Parise, H., Sullivan, L. & Meigs, J. B. Metabolic syndrome as a precursor of cardiovascular disease and type 2 diabetes mellitus. *Circulation* **112**, 3066–3072 (2005).
- Weindruch, R. & Walford, R. L. *The Retardation of Aging and Disease by Dietary Restriction* (Charles C. Thomas, Springfield, Illinois, 1988).
- Holliday, R. Food, reproduction, and longevity: is the extended life span of calorie-restricted animals and evolutionary adaptation? *BioEssays* **10**, 125–127 (1989).
- Kaeberlein, M., McVey, M. & Guarente, L. The SIR2/3/4 complex and SIR2 alone promote longevity in *Saccharomyces cerevisiae* by two different mechanisms. *Genes Dev.* **13**, 2570–2580 (1999).
- Tissenbaum, H. A. & Guarente, L. Increased dosage of a sir-2 gene extends lifespan in *Caenorhabditis elegans*. *Nature* **410**, 227–230 (2001).
- Wood, J. G. *et al.* Sirtuin activators mimic caloric restriction and delay ageing in metazoans. *Nature* **430**, 686–689 (2004).
- Imai, S., Armstrong, C. M., Kaeberlein, M. & Guarente, L. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795–800 (2000).
- Landry, J. *et al.* The silencing protein SIR2 and its homologs are NAD-dependent protein deacetylases. *Proc. Natl Acad. Sci. USA* **97**, 5807–5811 (2000).
- Fernandes, G., Yunis, E. J. & Good, R. A. Suppression of adenocarcinoma by the immunological consequences of calorie restriction. *Nature* **263**, 504–507 (1976).
- Zhu, H., Gou, Q. & Mattson, M. P. Dietary restriction protects hippocampal neurons against the death-promoting action of presenilin-1 mutation. *Brain Res.* **842**, 224–229 (1999).
- Ingram, D. K., Weindruch, R., Spangler, E. L., Freeman, J. R. & Walford, R. L. Dietary restriction benefits learning and motor performance of aged mice. *J. Gerontol.* **42**, 78–81 (1987).
- Sinclair, D. A. & Guarente, L. Extrachromosomal rDNA circles — a cause of aging in yeast. *Cell* **91**, 1033–1042 (1997).
- Aguilani, H., Gustafsson, L., Rigoulet, M. & Nystrom, T. Asymmetric inheritance of oxidatively damaged proteins during cytokinesis. *Science* **299**, 1751–1753 (2003).
- Lin, S. J., Defeseze, P. A. & Guarente, L. Requirement of NAD and SIR2 for life-span extension by calorie restriction in *Saccharomyces cerevisiae*. *Science* **289**, 2126–2128 (2000).
- Lin, S. J. *et al.* Calorie restriction extends *Saccharomyces cerevisiae* lifespan by increasing respiration. *Nature* **418**, 344–348 (2002).
- Lamming, D. W. *et al.* HST2 mediates SIR2-independent life-span extension by calorie restriction. *Science* **309**, 1861–1864 (2005).
- Lin, S.-J., Ford, E., Haigis, M., Liszt, G. & Guarente, L. Calorie restriction extends yeast life span by lowering the level of NADH. *Genes Dev.* **18**, 12–16 (2004).
- Anderson, R. M., Bitterman, K. J., Wood, J. G., Medvedik, O. & Sinclair, D. A. Nicotinamide and PNC1 govern lifespan extension by calorie restriction in *Saccharomyces cerevisiae*. *Nature* **423**, 181–185 (2003).
- Kaeberlein, M., Kirkland, K. T., Fields, S. & Kennedy, B. K. Sir2-independent life span extension by calorie restriction in yeast. *PLoS Biol.* **2**, e296 (2004).
- Kaeberlein, M. *et al.* Regulation of yeast replicative life span by TOR and Sch9 in response to nutrients. *Science* **310**, 1193–1197 (2005).
- Rogina, B., Helfand, S. L. & Frankel, S. Longevity regulation by *Drosophila* Rpd3 deacetylase and caloric restriction. *Science* **298**, 1745 (2002).
- Rogina, B. & Helfand, S. L. Sir2 mediates longevity in the fly through a pathway related to calorie restriction. *Proc. Natl Acad. Sci. USA* **101**, 15998–16003 (2004).
- Wang, Y. & Tissenbaum, H. A. Overlapping and distinct functions for a *Caenorhabditis elegans* SIR2 and DAF-16/FOXO. *Mech. Ageing Dev.* **127**, 48–56 (2006).
- Chen, D., Steele, A. D., Lindquist, S. & Guarente, L. Increase in activity during calorie restriction requires Sir1. *Science* **310**, 1641 (2005).
- Luo, J. *et al.* Negative control of p53 by Sir2a promotes cell survival under stress. *Cell* **107**, 137–148 (2001).
- Vaziri, H. *et al.* hSIRT2^{SIRT1} functions as an NAD-dependent p53 deacetylase. *Cell* **107**, 149–159 (2001).
- Motta, M. C. *et al.* Mammalian SIRT1 represses forkhead transcription factors. *Cell* **116**, 551–563 (2004).
- Brunet, A. *et al.* Stress-dependent regulation of FOXO transcription factors by the SIRT1 deacetylase. *Science* **303**, 2011–2015 (2004).
- Cohen, H. Y. *et al.* Acetylation of the C terminus of Ku70 by CBP and PCAF controls bax-mediated apoptosis. *Mol. Cell* **13**, 627–638 (2004).
- Cohen, H. Y. *et al.* Calorie restriction promotes mammalian cell survival by inducing the SIRT1 deacetylase. *Science* **305**, 390–392 (2004).
- Picard, F. *et al.* Sirt1 promotes fat mobilization in white adipocytes by repressing PPAR-γ. *Nature* **429**, 771–776 (2004).
- Bordone, L. *et al.* Sirt1 regulates insulin secretion by repressing UCP2 in pancreatic β cells. *PLoS Biol.* **4**, e31 (2005).
- Moynihan, K. A. *et al.* Increased dosage of mammalian Sir2 in pancreatic β cells enhances glucose-stimulated insulin secretion in mice. *Cell Metab.* **2**, 105–117 (2005).
- Kitamura, Y. I. *et al.* FoxO1 protects against pancreatic β cell failure through NeuroD and MafA induction. *Cell Metab.* **2**, 153–163 (2005).
- Rodgers, J. T. *et al.* Nutrient control of glucose homeostasis through a complex of PGC-1α and SIRT1. *Nature* **434**, 113–118 (2005).
- Nemoto, S., Fergusson, M. M. & Finkel, T. SIRT1 functionally interacts with the metabolic regulator and transcriptional coactivator PGC-1α. *J. Biol. Chem.* **280**, 16456–16460 (2005).
- Accili, D. & Arden, K. C. FoxOs at the crossroads of cellular metabolism, differentiation, and transformation. *Cell* **117**, 421–426 (2004).
- Taniguchi, C. M., Emanuelli, B. & Kahn, C. R. Critical nodes in signaling pathways: insights into insulin action. *Nature Rev. Mol. Cell Biol.* **7**, 85–96 (2006).
- Nisoli, E. *et al.* Calorie restriction promotes mitochondrial biogenesis by inducing the expression of eNOS. *Science* **310**, 314–317 (2005).
- Hagopian, K., Ramsey, J. J. & Weindruch, R. Influence of age and calorie restriction on liver glycolytic enzyme activities and metabolite concentrations in mice. *Exp. Gerontol.* **38**, 253–266 (2003).
- Wallace, D. C. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu. Rev. Genet.* **39**, 359–407 (2005).
- Onyango, P., Celic, I., McCaffery, J. M., Boeke, J. D. & Feinberg, A. P. SIRT3, a human SIR2 homologue, is an NAD-dependent deacetylase localized to mitochondria. *Proc. Natl Acad. Sci. USA* **99**, 13653–13658 (2002).
- Schwer, B., North, B. J., Frye, R. A., Ott, M. & Verdin, E. The human silent information regulator (Sir2) homologue hSIRT3 is a mitochondrial nicotinamide adenine dinucleotide-dependent deacetylase. *J. Cell Biol.* **158**, 647–657 (2002).
- Haigis, M. C. *et al.* SIRT4 inhibits glutamate dehydrogenase and opposes the effects of calorie restriction in pancreatic β-cells. *Cell* **126**, 941–956 (2006).
- Schwer, B., Bunkenborg, J., Verdin, R. O., Andersen, J. S. & Verdin, E. Reversible lysine acetylation controls the activity of the mitochondrial enzyme acetyl-CoA synthetase2. *Proc. Natl Acad. Sci. USA* **103**, 10224–10229 (2006).
- Hallows, W. C., Lee, S. & Denu, J. M. Sirtuins deacetylate and activate mammalian acetyl-CoA synthetases. *Proc. Natl Acad. Sci. USA* **103**, 10230–10235 (2006).
- Starai, V. J., Celic, I., Cole, R. N., Boeke, J. D. & Escalante-Semerena, J. C. Sir2-dependent activation of acetyl-CoA synthetase by deacetylation of active lysine. *Science* **298**, 2390 (2002).
- Buckley, B. M. & Williamson, D. H. Origins of blood acetate in the rat. *Biochem. J.* **166**, 539–545 (1977).
- Fujino, T., Kondo, J., Ishikawa, M., Morikawa, K. & Yamamoto, T. Acetyl-CoA synthetase 2, a mitochondrial matrix enzyme involved in the oxidation of acetate. *J. Biol. Chem.* **276**, 11420–11426 (2001).
- Mostoslavsky, R. *et al.* Genomic instability and aging-like phenotype in the absence of mammalian SIRT6. *Cell* **124**, 315–329 (2006).
- Michishita, E., Park, J. Y., Burneski, J. M., Barrett, J. C. & Horikawa, I. Evolutionarily conserved and nonconserved cellular localizations and functions of human SIRT proteins. *Mol. Biol. Cell* **16**, 4623–4635 (2005).
- Ford, E., Voit, R., Liszt, G., Grummt, I. & Guarente, L. Mammalian Sir2 homolog SIRT7 is an activator of RNA polymerase I transcription. *Genes Dev.* **20**, 1075–1081 (2006).

55. Muth, V., Nadaud, S., Grummt, I. & Voit, R. Acetylation of TAF₆₈, a subunit of TFIIB/SLI, activates RNA polymerase I transcription. *EMBO J.* **20**, 1353–1362 (2001).
56. Puigserver, P. *et al.* A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell* **92**, 829–839 (1998).
57. Lin, J., Handschin, C. & Spiegelman, B. M. Metabolic control through the PGC-1 family of transcription coactivators. *Cell Metab.* **1**, 361–370 (2005).
58. Mootha, V. K. *et al.* PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet.* **34**, 267–273 (2003).
59. Patti, M. E. *et al.* Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: potential role of PGC1 and NRF1. *Proc. Natl Acad. Sci. USA* **100**, 8466–8471 (2003).
60. Puigserver, P. *et al.* Insulin-regulated hepatic gluconeogenesis through FOXO1-PGC-1 α interaction. *Nature* **423**, 550–555 (2003).
61. Lin, J. *et al.* Hyperlipidemic effects of dietary saturated fats mediated through PGC-1 β coactivation of SREBP. *Cell* **120**, 261–273 (2005).
62. Wolfrum, C. & Stoffel, M. Coactivation of Foxa2 through Pgc-1 β promotes liver fatty acid oxidation and triglyceride/VLDL secretion. *Cell Metab.* **3**, 99–110 (2006).
63. Wolfrum, C., Asilmaz, E., Luca, E., Friedman, J. & Stoffel, M. Foxa2 regulates lipid metabolism and ketogenesis in the liver during fasting and in diabetes. *Nature* **432**, 1027–1032 (2004).
64. Kahn, B. K., Alquier, T., Carling, D. & Hardie, D. G. AMP-activated protein kinase: ancient energy gauge provides clues to modern understanding of metabolism. *Cell Metab.* **1**, 15–25 (2005).
65. Hardie, D. G., Scott, J. W., Pan, D. A. & Hudson, E. R. Management of cellular energy by the AMP-activated protein kinase system. *FEBS Lett.* **546**, 113–120 (2003).
66. Hawley, S. A. *et al.* Complexes between the LKB1 tumor suppressor, STRAD α/β and Mo25 α/β are upstream kinases in the AMP-activated protein kinase cascade. *J. Biol.* **2**, 1–16 (2003).
67. Woods, A. *et al.* LKB1 is the upstream kinase in the AMP-activated protein kinase cascade. *Curr. Biol.* **13**, 2004–2008 (2003).
68. Shaw, R. J. *et al.* The tumor suppressor LKB1 kinase directly activates AMP-activated kinase and regulates apoptosis in response to energy stress. *Proc. Natl Acad. Sci. USA* **101**, 3329–3335 (2004).
69. Birnbaum, M. J. Activating AMP-activated protein kinase without AMP. *Mol. Cell* **19**, 289–296 (2005).
70. Zhou, G. *et al.* Role of AMP-activated protein kinase in mechanism of metformin action. *J. Clin. Invest.* **108**, 1167–1174 (2001).
71. Brownsey, R. W., Boone, A. N., Elliott, J. E., Kulpa, J. E. & Lee, W. M. Regulation of acetyl-CoA carboxylase. *Biochem. Soc. Trans.* **34**, 223–227 (2006).
72. Koo, S.-H. *et al.* The CREB coactivator TORC2 is a key regulator of fasting glucose metabolism. *Nature* **437**, 1109–1114 (2005).
73. Canettieri, G. *et al.* Dual role of the coactivator TORC2 in modulating hepatic glucose output and insulin signaling. *Cell Metab.* **2**, 331–338 (2005).
74. Shaw, R. J. *et al.* The kinase LKB1 mediates glucose homeostasis in liver and therapeutic effects of metformin. *Science* **310**, 1642–1646 (2005).
75. Bauer, J. A. *et al.* Resveratrol improves health and survival of mice on a high-calorie diet. *Nature* **444**, 337–342 (2006).
76. Lagouge, M. *et al.* Resveratrol improves mitochondrial function and protects against metabolic disease by activating SIRT1 and PGC-1 α . *Cell* **127**, 1–14 (2006).

Acknowledgements The author apologizes for the many studies and references that could not be included because of space limitations. Work from the author's laboratory was supported by the NIH.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to the author (leng@mit.edu).

Mechanisms linking obesity with cardiovascular disease

Luc F. Van Gaal¹, Ilse L. Mertens¹ & Christophe E. De Block¹

Obesity increases the risk of cardiovascular disease and premature death. Adipose tissue releases a large number of bioactive mediators that influence not only body weight homeostasis but also insulin resistance — the core feature of type 2 diabetes — as well as alterations in lipids, blood pressure, coagulation, fibrinolysis and inflammation, leading to endothelial dysfunction and atherosclerosis. We are now beginning to understand the underlying mechanisms as well as the ways in which smoking and dyslipidaemia increase, and physical activity attenuates, the adverse effects of obesity on cardiovascular health.

Obesity is a fast growing problem that is reaching epidemic proportions worldwide¹, and is associated with an increased risk of premature death². Individuals with a central deposition of adipose tissue can experience elevated cardiovascular morbidity and mortality, including stroke, congestive heart failure, myocardial infarction and cardiovascular death, and this is independent of the association between obesity and other cardiovascular risk factors^{3,4}. Weight gain during adult life⁵ or even during childhood and adolescence⁶ seems to have a great effect on diabetes and cardiovascular risk, even within the normal body mass index (BMI) range.

Owing to the increasing rates of childhood obesity, the global life expectancy in the United States will, for the first time in recent history, decline⁷, and the American Heart Association (AHA) has reclassified obesity as a 'major, modifiable risk factor' for coronary heart disease (CHD)⁸.

Different mechanisms linking obesity to cardiovascular disease have been postulated, adding weight to classical risk factors. Obesity also increases the prevalence of less conventional risk factors. Adipose tissue is an active endocrine and paracrine organ that releases a large number of cytokines and bioactive mediators, such as leptin, adiponectin, interleukin-6 (IL-6) and tumour necrosis factor- α (TNF- α), that influence not only body weight homeostasis but also insulin resistance, diabetes, lipid levels, tension, coagulation, fibrinolysis, inflammation and atherosclerosis⁹. Various morphological adaptations in cardiac structure and haemodynamic function also occur in obese individuals¹⁰.

Here we describe and discuss classical and less conventional risk factors within the scope of insulin resistance as a major contributing factor to the obesity-associated cardiovascular risk.

Risk factors

Classical cardiovascular risk factors include smoking, high low-density lipoprotein (LDL) cholesterol, hypertension and dysfunctions in glucose metabolism. The extent to which these factors might contribute to an independent, additional risk among obese individuals is unclear. When compared with non-obese smokers, obese smokers have a clear reduction in life expectancy^{11,12}, so smoking clearly confounds considerations based on BMI alone.

Obesity induces several cytokines and inflammatory markers that might contribute to the cardiovascular outcome in overweight and obese people. It is also associated with increased levels of endothelial

cell products such as intercellular adhesion molecule-1 (ICAM-1)¹³. Cigarette smoking may also influence cytokines: it causes a dose-dependent increase in plasma ICAM-1 (ref. 14) and decrease in adiponectin levels^{15,16}, thereby inducing endothelial dysfunction. Whether these effects are direct effects of nicotine, carbon monoxide or other components of tobacco smoke is unclear. Both hydrogen peroxide and nicotine suppress adiponectin expression in adipocytes¹⁵. Furthermore, smoking may directly regulate adiponectin concentration through lipolysis. Nicotinic acetylcholine receptors may be involved in the regulation of adipokine (also known as adipocytokine) release in adipocytes *in vitro*¹⁷.

Smoking may also increase levels of tumour necrosis factor- α (TNF- α), a powerful cytokine that might decrease adiponectin levels and induce insulin resistance. Lower adiponectin levels found in chronic smokers confirm earlier findings that chronic smokers are insulin resistant¹⁸. Cigarette smoking may also enhance the oxidative stress (production of reactive oxygen species, ROS) resulting from obesity owing to the direct effects of the free radicals present in smoke, as well as the inflammatory response it can induce¹⁹. Endothelial dysfunction, an early marker of atherosclerosis and present early in obesity, has also been observed in chronic cigarette smokers¹⁴. This could occur either directly or through the effect of smoking on interleukin-6 (IL-6) secretion from adipocytes and subsequent C-reactive protein (CRP) burst. Figure 1 depicts how smoking might have additional negative effects in obese individuals.

Dyslipidaemia in obesity is characterized by increased levels of very low-density lipoprotein (VLDL) cholesterol, triacylglycerols and total cholesterol, an increase in small dense LDL particles, and lower high-density lipoprotein (HDL) cholesterol levels²⁰. As to the effect of lipoproteins among conventional risk factors, there is little evidence that LDL cholesterol — known to be a major risk factor for CHD — is enhanced in obesity, among patients with visceral obesity in particular. Other lipoproteins, such as the apolipoprotein (Apo) B/A1 ratio²¹, small dense LDL particles and low HDL cholesterol, are predictive for cardiovascular disease endpoints in people with visceral adiposity.

It is not clear whether hypercholesterolaemia further increases the risk of cardiovascular disease in obese individuals. But the insulin-resistant state of abdominal obesity adds substantially to the CHD risk of patients with familial hypercholesterolaemia²². Obesity also limits the beneficial effects of lipid-lowering strategies²³. However, aggressive reduction of lipids with the use of statins can have a beneficial effect on coronary plaque development in obese individuals²³.

¹University of Antwerp, Faculty of Medicine, Department of Diabetology, Metabolism and Clinical Nutrition, Antwerp University Hospital, Wilrijkstraat 10, B-2650 Edegem (Antwerp), Belgium.

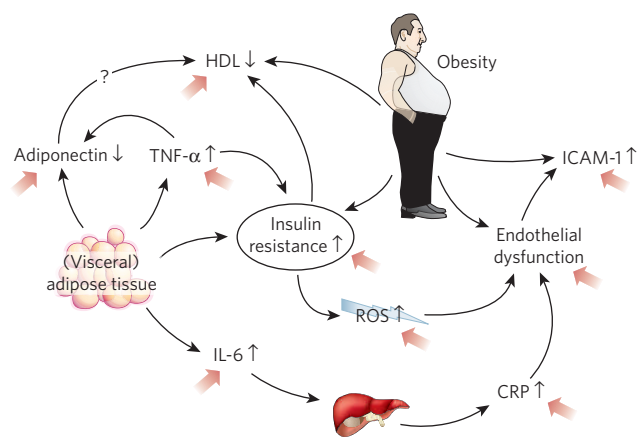


Figure 1 | Schematic representation of how smoking might add to several mechanisms linking obesity to cardiovascular disease. Both abdominal obesity and smoking are associated with insulin resistance, oxidative stress and increased levels of different (adipo)cytokines and inflammatory markers, all of which ultimately lead to endothelial dysfunction. Red arrows indicate an effect of smoking.

In insulin-resistant states, the dyslipidaemia as seen in obesity is characterized by a different composition and distribution of LDL particles, resulting in an increased concentration of small, dense LDL. LDL particles are enriched in triacylglycerols, which are rapidly lipolysed by hepatic lipase (HL) leaving smaller, denser LDL particles. The activities of both cholesterol ester transfer protein (CETP) and HL seem to be increased in metabolic syndrome, leading to such particles, which are more prone to oxidation and glycation. Small dense LDL particles can move through endothelial fenestrations, entering the subendothelial space where inflammation and transformation into plaque can occur²⁴. The modified LDL is mostly taken up by macrophage scavenger receptors, rather than the normal LDL receptor (LDLR) pathway, thus inducing atherosclerosis.

Hepatic overproduction of VLDL seems to be the primary and crucial defect in obesity — a consequence of hepatic steatosis. Insulin resistance in the liver, muscles and adipose tissue leads to an inability to suppress hepatic glucose production, impaired glucose uptake and oxidation, and inability to suppress release of non-esterified fatty acids (NEFAs) from adipose tissue. Another key factor in the regulation of VLDL secretion is the rate of ApoB-100 degradation, which is decreased in insulin resistance. In addition to increased synthesis, there is also decreased clearance of triacylglycerol-rich lipoproteins (TRLs) induced by a decrease in lipoprotein lipase activity²⁵. The insulin resistance associated with obesity might also impair LDLR activity, thus contributing to the delayed VLDL-particle clearance²⁶. Plasma ApoC-III concentration functions as an inhibitor of lipoprotein lipase, contributing to the kinetic defects in ApoB metabolism²⁷. This may explain its strong association with triacylglycerols and the fact that it is a predictor of coronary artery disease.

Several mechanisms contribute to the decreased HDL seen in obesity. Impaired lipolysis of TRLs leads to reduced HDL concentration by decreasing the transfer of apolipoproteins and phospholipids from TRL to HDL. In addition, the delayed clearance of TRLs facilitates the CETP-mediated exchange between cholesterol esters in HDL and triacylglycerols in VLDL. The increased activity of hepatic lipase in insulin-resistant states such as obesity produces smaller HDL particles, leading to increased HDL elimination²⁸. Insulin resistance could also have a direct effect on the production of ApoA-I or hepatic secretion of nascent HDL. This insulin resistance may be the triggering factor of metabolic syndrome as a cluster, with insulin resistance and abdominal obesity as core features but probably also associated with other conditions such as physical inactivity, proinflammatory and pro-coagulation states²⁹, and oxidation.

Fitness versus fatness

Most obesity studies have not adequately measured physical activity and functional capacity, which are known to predict risk for CHD, and the independent contributions of 'fitness' versus 'fatness' to cardiovascular risk are still debated³⁰.

Both increased adiposity and reduced physical activity are strong and independent predictors of CHD³¹ and death^{5,12,32,33}. In general, for each unit of BMI increment, the risk of CHD increases by 8%³¹. On the other hand, each 1 h-MET (metabolic equivalent) increase in activity score is associated with an 8% decrease in CHD risk. Most heart attacks can be predicted from nine easily measurable and modifiable factors, including abdominal adiposity and regular physical activity²¹.

In most studies^{5,12,31,32}, being physically active moderately attenuated but did not eliminate the adverse effect of obesity on coronary health, and being lean did not counteract the increased risk associated with physical inactivity. But several studies lend support to current exercise guidelines, which endorse moderate-intensity exercise for at least 30 minutes per day³¹. Weight gain during adulthood is a strong and independent risk factor for premature death, regardless of the level of physical activity⁵.

Physical activity improves glucose tolerance and sensitivity by improving non-insulin-dependent glucose uptake; it improves the ratio between HDL and LDL cholesterol because it increases the activity of lipoprotein lipase; it decreases triacylglycerols, increases fibrinolysis, decreases platelet aggregation, improves oxygen uptake in the heart as well as in peripheral tissues, lowers the resting heart rate by increasing vagal tone, and lowers blood pressure³³. Physical activity also directly increases myocardial oxygen supply, improving myocardial contraction and electrical stability.

Insulin resistance

Insulin resistance often clusters with various classical cardiovascular risk factors such as lipid abnormalities, glucose intolerance and high blood pressure. Whether insulin resistance is involved in the atherogenic process and manifests itself in the heart and the vasculature is controversial³⁴. Normal insulin signalling in cardiovascular tissues is mediated by two different pathways: one that is predominant in metabolic tissues (the phosphatidylinositol-3-OH kinase pathway) and a growth-factor-like pathway (mediated by mitogen-activated protein kinase). In cardiovascular tissues, insulin resistance leads to inhibition of the metabolic pathway and to overstimulation of the growth-factor-like pathway³⁴, and might lead to a decrease in glucose uptake, possibly hampering normal cardiac function³⁵.

Several of the possible mechanisms linking obesity to cardiovascular disease, such as increased levels of NEFAs, lipotoxicity and disturbances in adipokine secretion, are believed to be related to insulin resistance. In people with normal insulin sensitivity, lipolysis of adipose tissue is well regulated with the release of NEFAs relative to the energy requirement of the different tissues. Insulin resistance in hypertrophic adipocytes leads to increased lipolysis and the release of NEFAs. But other mechanisms could also lead to increased NEFA levels, such as decreased fatty acid oxidation and low levels of adiponectin, which normally activates fatty acid oxidation by turning on the fuel sensor AMP-activated protein kinase. Other factors involved in NEFA behaviour include stress-induced increases in adrenergic tone and the inflammatory state associated with obesity. Increased levels of NEFAs might affect endothelial nitric oxide production, thereby impairing endothelium-dependent vasodilation. They may also increase myocardial oxygen requirements — and therefore ischaemia — decrease myocardial contractility and induce cardiac arrhythmias³⁶. High NEFA concentrations may increase ROS generation in mononuclear cells, and induce insulin resistance in myocytes and hepatocytes. Recent evidence in older men with coronary artery disease has also shown that NEFAs are independently associated with cardiovascular mortality³⁷.

Several adipokines, such as leptin, adiponectin, TNF- α , IL-6, resistin and the recently discovered visfatin and retinol-binding protein 4 (RBP4)³⁸, have been suggested to be associated with insulin resistance.

RBP4 is an adipocyte-secreted molecule that is elevated in the serum before the development of diabetes and seems to signal the presence of insulin resistance and associated cardiovascular risk factors³⁸. RBP4 seems to be involved in impairing insulin signalling in muscles and inducing the expression of gluconeogenic enzymes in the liver, contributing to insulin resistance³⁹.

When adipocytes exceed their storage capacity and the process of fat-cell proliferation fails, fat begins to accumulate in tissues not suited for lipid storage, leading to the formation of specific metabolites that inhibit insulin signal transduction⁴⁰.

Visceral and ectopic fat

In addition to total body fatness, the accumulation of abdominal fat independently increases cardiovascular risk. The waist-to-hip ratio reflects abdominal fat in predicting type 2 diabetes, stroke, myocardial infarction and cardiovascular mortality in middle-aged individuals. The Nurses' Health Study confirmed these findings using the waist circumference⁴¹. Moreover, visceral fat correlates with cardiovascular risk factors⁴². There is an independent curvilinear association between visceral adiposity and mortality⁴³, suggesting that a large amount of visceral fat is required for an increased risk of mortality.

Insulin resistance may be one of the most important factors linking abdominal visceral adiposity to cardiovascular risk. Impaired suppression of adipocyte lipolysis and elevated NEFA levels are also associated with abdominal adiposity, potentially leading to vascular endothelial dysfunction. Visceral adiposity is also associated with increased levels of the pro-coagulant plasminogen activator inhibitor-1 (PAI-1) and low-grade inflammation. Besides these metabolic disorders, visceral fat accumulation is correlated with systolic blood pressure in postmenopausal women, and visceral fat reduction was directly associated with lowering of blood pressure.

Visceral fat accumulation is associated not only with quantitative changes in serum lipids but also with qualitative changes in lipoproteins, such as small dense LDL, and conveys greater insulin resistance than subcutaneous fat.

A number of adipokines might originate from non-adipose cells in adipose tissue — macrophages in particular — and these include atherogenic cytokines⁴⁴. Different factors could induce macrophage infiltration and activation in white adipose tissue such as adipocyte hypertrophy and hyperplasia, secretion of monocyte chemoattractant protein-1 (MCP-1) or local actions of leptin or adiponectin⁴⁵. Macrophages are found more frequently in visceral compared with subcutaneous adipose tissue⁴⁶. The mechanisms by which macrophages are recruited presumably involve secretion of chemotactic molecules by adipose tissue.

To what extent visceral fat exerts a direct effect on risk, of mortality in particular, or indirect effects, through insulin resistance or the effects of adipokines, remains an open question. The emerging evidence of excess release of proinflammatory and prothrombotic factors seems to be gaining more weight as the underlying mechanism than the originally proposed portal/fatty acid flux theory⁴⁷. It is not yet clear whether visceral fat should be considered to be a variant of ectopic fat, such as liver or muscle fat⁴⁸, or an incapacity of the body to store fat in pre-designed subcutaneous and/or gluteal areas. Liver fat has been directly associated with mortality risk⁴³. This idea about fat topography is in line with findings that peripheral, particularly gluteal, fat might even have a protective effect⁴⁹.

Ectopic fat storage in the heart, blood vessels and kidneys can impair their function, contributing to the increased cardiovascular risk in obesity⁵⁰. Besides the cardiac alterations that result from haemodynamic changes and hypertension, excessive lipid accumulation in the myocardium induced by weight gain may be directly cardiotoxic. In animal models, cardiac lipotoxicity causes eccentric left ventricular remodeling, increased left ventricular pressure and decreased systolic performance^{51,52}, leading, ultimately, to dilated cardiomyopathy. Accumulation of large amounts of myolipids may result in apoptosis and systolic dysfunction. Fatty acid transfer protein-1 expression in the heart, leading to lipotoxic cardiomyopathy, may also contribute to obesity-induced

increased myocardial oxygen consumption. In humans, myocardial triacylglycerol content increases with increasing BMI⁵³.

In peripheral vessels, high amounts of perivascular fat cells could contribute mechanically to the increased vascular stiffness seen in obesity. Periadventitial adipose tissue in particular may regulate the arterial tone of mesenteric arteries⁵⁴, with increased arterial stiffness as a consequence. In addition, increased adipose paracrine secretion may lead to vascular smooth muscle cell (VSMC) growth induced by growth factors. This is certain to occur in the absence of adiponectin, which is known to reduce VSMC proliferation and migration⁵⁵.

Adipokines

Adipose tissue is no longer viewed as a passive organ for triacylglycerol storage and as a source of NEFAs. As developing preadipocytes differentiate to become mature adipocytes, they acquire the ability to secrete various proteins, many of which are released as cytokines. Mature adipocytes act as an active endocrine and paracrine organ, secreting an ever-increasing number of growth factors that participate in diverse metabolic processes, particularly insulin resistance. Compounds that influence different metabolic and vascular processes, including adipogenesis, are lipoprotein lipase, CETP, angiotensinogen, complement factors (adipsin or complement D), adiponectin, acylation-stimulating protein (ASP), IL-6, prostaglandins, TNF- α , a newly identified protein/cytokine known as adipocyte differentiation factor and, as recently discovered, nitric oxide¹⁶.

After the discovery of leptin as a specific adipose-tissue-derived hormone, other adipokines have been found to contribute to insulin resistance, metabolic disturbances and cardiovascular risk. These include adiponectin, TNF- α , PAI-1 and, recently, visfatin.

Leptin is a bioactive substance found in adipose tissue that controls food intake and energy expenditure. It also has atherogenic and growth properties. Human obesity is associated with elevated leptin levels, which have been proposed to have a role in insulin resistance and metabolic syndrome. There may be a direct link between hyperleptinaemia and increased cardiovascular disease risk. Leptin may enhance platelet aggregation and arterial thrombosis⁵⁶, promote angiogenesis, impair arterial distensibility and induce proliferation and migration of VSMCs⁵⁷. In view of the leptin-induced thrombogenicity among obese men and women, increased leptin levels are related to PAI-1 (both sexes) and von Willebrand factor (men only), independently of fatness, insulin resistance and amount of visceral fat⁵⁸. In addition to these mechanisms,

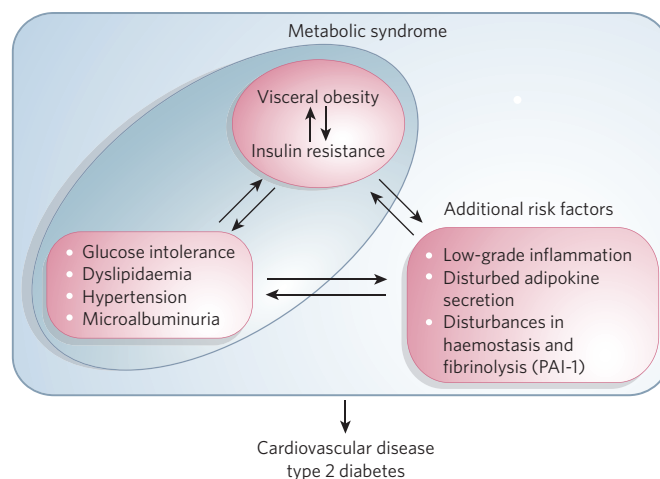


Figure 2 | Disturbances in haemostasis and fibrinolysis as additional cardiovascular risk factors in relation to metabolic syndrome. Levels of PAI-1, a marker of hypofibrinolysis, are increased in individuals with metabolic syndrome and are associated with both classical features and new components of metabolic syndrome and its core features insulin resistance and visceral obesity.

leptin seems to enhance the calcification of vascular cells, making the arterial wall a target of leptin.

Adiponectin has important anti-atherogenic, antidiabetic and anti-inflammatory properties, and is expressed abundantly in adipocytes^{9,16}. Unlike most other adipokines, adiponectin is decreased in obesity, diabetes and other insulin-resistant states. The mechanism by which plasma levels are reduced in individuals with visceral fat accumulation has not yet been clarified, but an increase in TNF- α secretion from accumulated visceral fat seems to have a role. Co-culture with visceral fat cells inhibits adiponectin secretion from subcutaneous adipocytes, suggesting that some inhibiting factors for adiponectin synthesis or secretion are secreted from visceral adipose tissue. The negative correlation between visceral adiposity and adiponectin levels might be explained by the increased secretion of TNF- α from accumulated visceral fat.

Adiponectin increases the expression of messenger RNA and protein production of tissue inhibitor of metalloproteinase in macrophages through the induction of IL-10 synthesis⁵⁹, and selectively suppresses endothelial cell apoptosis⁶⁰. This suggests that adiponectin protects plaque rupture by the inhibition of matrix metalloproteinase function. Endothelium-dependent vasoreactivity is impaired in people with hypoadiponectinaemia, which might be at least one cause of hypertension in visceral obesity⁶¹. The protein inhibits the expression of adhesion molecules, such as vascular cell adhesion molecule-1 (VCAM-1), ICAM-1 and E-selectin, through the inhibition of nuclear factor- κ B (NF- κ B) activation; it also suppresses foam-cell formation.

Visfatin is a visceral-fat-specific protein that is probably involved in the development of obesity-related diseases. Visfatin levels correlate strongly with an individual's amount of visceral fat⁶², exert insulin-mimetic effects in culture cells, and have a potent insulin-like activity of adipogenesis. Because of visfatin's possible — but already controversial⁶³ — origin in visceral adipose tissue, it is thought that future research might reveal an independent role of visceral fat in cardiovascular diseases.

Pro-coagulation and hypofibrinolysis

The prothrombotic state in the atherosclerotic process encompasses platelet hyperaggregability, hypercoagulability and hypofibrinolysis. In obesity and metabolic syndrome, fibrinogen, von Willebrand factor (vWF) and PAI-1 have been most extensively studied as markers of the haemostatic and fibrinolytic system and as possible predictors for the development of cardiovascular disease and type 2 diabetes⁶⁴. In obese individuals, only PAI-1 levels were increased in those with metabolic syndrome²⁹ (Fig. 2). PAI-1 is expressed in visceral adipose tissue. It is mainly expressed in stromal cells, including monocytes, smooth muscle cells and pre-adipocytes⁶⁵. Visceral adipose tissue seems to have up to five times the number of PAI-1-producing stromal cells compared with subcutaneous adipose tissue. Plasma PAI-1 levels are more closely related to fat accumulation and PAI-1 expression in the liver than in adipose tissue⁶⁶, suggesting that in insulin-resistant individuals the fatty liver is an important site of PAI-1 production. This confirms the central role of the liver in these processes. The fibrinolytic system could have a role in the regulation of adipose tissue development and insulin signalling in adipocytes⁶⁷.

Fibrinogen, vWF and PAI-1 are also considered to be markers of the acute-phase reaction of inflammation, and thrombosis has been closely linked to inflammation⁶⁸. Fibrinogen clusters with inflammation variables rather than with markers of pro-coagulation or metabolic factors⁶⁹. CRP increases PAI-1 expression and activity in human aortic endothelial cells⁷⁰. Figure 2 shows how disturbances in haemostasis and fibrinolysis relate to metabolic syndrome and its different components.

Oxidative stress

Oxidative stress has been proposed to be a potential pathogenic mechanism linking obesity and insulin resistance with endothelial dysfunction⁷¹. However, it is not clear whether obesity itself or obesity-associated conditions lead to oxidative stress. Several pathways

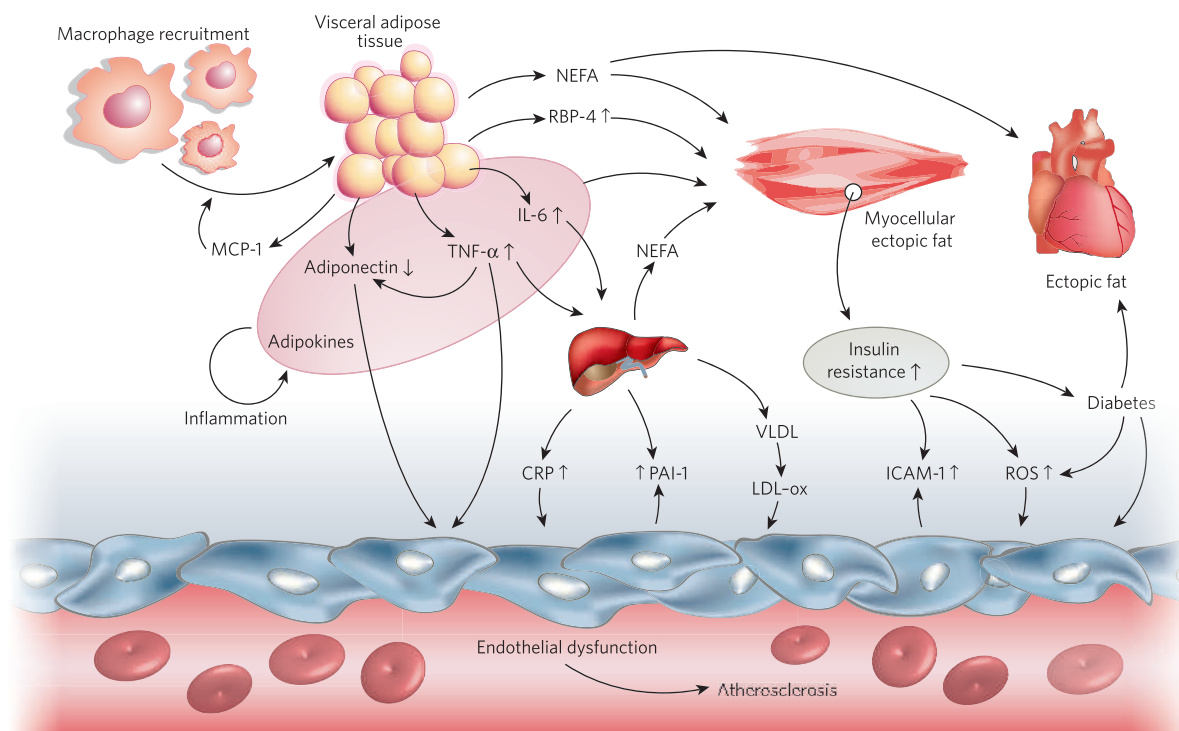


Figure 3 | Both abdominal (visceral) fat and insulin resistance may contribute to cardiovascular disease in obesity. Visceral fat in particular contributes to endothelial dysfunction through the direct effect of adipokines, mainly adiponectin and TNF- α , which are secreted by fat tissue after macrophage recruitment (through monocyte chemoattractant protein-1, MCP-1). Indirect effects of TNF- α and IL-6 might influence

inflammation (CRP) and endothelial dysfunction. Insulin resistance induced by cytokines (IL-6, TNF- α and adiponectin), NEFA and retinol-binding protein 4 (RBP-4) may induce oxidative stress and subsequent endothelial dysfunction (PAI-1 and ICAM-1). Fat accumulation, insulin resistance, liver-induced inflammation and dyslipidaemic features may all lead to the premature atherosclerotic process.

Box 1 | Insulin resistance and left ventricle remodelling

Potential mechanisms by which insulin resistance and its correlates might be involved in left ventricle remodelling and contractile dysfunction (data from refs 80, 81).

Increased cardiomyocyte lipid accumulation
Induction of ROS and apoptosis
Generation of non-enzymatic advanced glycation end products
Altered myocardial protein degradation
Insulin-like growth factor-mediated effects
Altered matrix remodelling
Altered vascular compliance
Sympathetic activation
Increased renal sodium reabsorption

are likely to contribute; excess weight and obesity are associated not only with increased oxidative stress, but also with elevated systemic inflammation, activation of the coagulation cascade, disturbances in the renin-angiotensin system, and, most importantly, enhanced lipid and protein oxidation resulting in the generation of oxidized LDL. Several insulin-resistant syndrome components predict high levels of oxidized LDL. Elevated oxidized LDL has been shown to predict myocardial infarction in well-functioning elderly people, even after adjusting for age, gender, race, smoking and metabolic syndrome⁷². Decrease in oxidative stress after dietary restriction and weight loss has been reported in obese individuals.

When caloric intake exceeds energy expenditure, the substrate-induced increase in tricarboxylic acid cycle activity generates an excess of mitochondrial NADH and ROS⁷³. It is reasonable to assume that processes that occur in muscle and fat cells might also occur in other cell types, particularly endothelial cells. Oxidative stress is considered to be the common factor underlying insulin resistance, type 2 diabetes and cardiovascular disease, and may explain the presence of inflammation in all of these conditions. The release of IL-6, mainly from abdominal adipocyte sources, might have a pivotal role in the relationship between oxidative stress and endothelial dysfunction. IL-6 not only contributes to CRP elevation and low-grade inflammatory state, but also has a close relationship with coagulation, insulin resistance, dyslipidaemia and endothelial dysfunction. Inflammatory cytokines are also acutely increased by hyperglycaemia⁷⁴ and NEFAs⁷⁵. Elevation of CRP enhances the CHD prognosis in men with four or five features of metabolic syndrome⁷⁶.

Endothelial dysfunction

It is generally assumed that traditional risk factors and insulin resistance are largely responsible for endothelial dysfunction and atherosclerosis. But inflammation might have an essential role in the development of insulin resistance, the initiation and progression of atherosclerotic lesions, and plaque disruption. IL-6 and TNF- α are inflammatory cytokines and the main inducers of CRP secretion in the liver. CRP is a marker of low-grade inflammation, and studies suggest that this protein has a role in the pathogenesis of atherosclerotic lesions in humans⁷⁷. Obesity-associated insulin resistance and low-grade inflammation induced by fat-cell-derived cytokines might help to explain other alterations in endothelial function as well as the clinically observed residual risk among such patients.

Obesity is associated with both increased local adipose and more general inflammation⁷⁸. Both adipocytes and resident macrophages may generate local adipose inflammatory output. They each synergistically stimulate inflammatory activity of the other⁷⁸. This autocrine and paracrine effect of IL-6 and TNF- α secretion is particularly pronounced in obesity⁷⁹, confirming adipose tissue to be a proinflammatory organ. Obesity is also associated with more generalized, systemic inflammation. This systemic inflammation, involving circulating inflammatory proteins such as CRP, IL-6, PAI-1, P-selectin, VCAM-1 and fibrinogen, is undoubtedly not only of adipocyte but also of hepatic origin.

Visceral and ectopic fat, its fatty acid behaviour, local adipokine secretion and low-grade inflammation in the context of an insulin-resistant obese patient, might explain the development of endothelial dysfunction and early cardiovascular disease.

Conclusion

Visceral obesity and insulin resistance increase cardiovascular risk by classical (dyslipidaemia, hypertension and glucose dysmetabolism) and less conventional mechanisms. Less conventional risk factors secreted by adipocytes and macrophages infiltrating adipose tissue, which is now considered to be an active endocrine organ, include adipokines (such as leptin and adiponectin), proinflammatory cytokines (IL-6 and CRP) and hypofibrinolytic factors (PAI-1) that, together, might lead to increased oxidative stress and endothelial dysfunction, finally promoting atherosclerosis (Fig. 3).

Common pathways involved in the pathogenesis of obesity and cardiovascular disease include insulin resistance and low-grade inflammation. The insulin-resistant state of obesity is also characterized by increased levels of NEFAs, which cause lipotoxicity and thereby impair endothelium-dependent vasodilation, increase oxidative stress and have cardiotoxic effects. Potential mechanisms linking central adiposity and insulin resistance with left ventricular remodelling⁸⁰ are summarized in Box 1.

When classical risk factors, including smoking, are superimposed on the insulin-resistant state of obesity they may further increase less conventional risk factors, exacerbating existing cardiovascular problems. ■

- James, P. T., Rigby, N. & Leach, R. International Obesity Task Force. The obesity epidemic, metabolic syndrome and future prevention strategies. *Eur. J. Cardiovasc. Prev. Rehabil.* **11**, 3–8 (2004).
- Adams, K. F. *et al.* Overweight, obesity, and mortality in a large prospective cohort of persons 50 to 71 years old. *N. Engl. J. Med.* **355**, 763–778 (2006).
- Lakka, T. A., Lakka, H. M., Salonen, R., Kaplan, G. A. & Salonen, J. T. Abdominal obesity is associated with accelerated progression of carotid atherosclerosis in men. *Atherosclerosis* **154**, 497–504 (2001).
- Kenchaiah, S. *et al.* Obesity and the risk of heart failure. *N. Engl. J. Med.* **347**, 305–313 (2002).
- Hu, F. B. *et al.* Adiposity as compared with physical activity in predicting mortality among women. *N. Engl. J. Med.* **351**, 2694–2703 (2004).
- Berenson, G. S. *et al.* Association between multiple cardiovascular risk factors and atherosclerosis in children and young adults. The Bogalusa Heart Study. *N. Engl. J. Med.* **338**, 1650–1656 (1998).
- Olshansky, S. J. *et al.* A potential decline in life expectancy in the United States in the 21st century. *N. Engl. J. Med.* **352**, 1138–1145 (2005).
- Eckel, R. H., Kahn, R., Robertson, R. M. & Rizza, R. A. Preventing cardiovascular disease and diabetes. A call to action from the American Diabetes Association and the American Heart Association. *Circulation* **113**, 2943–2946 (2006).
- Lau, D. C. W., Dhillon, B., Yan, H., Szmitko, P. E. & Verma, S. Adipokines: molecular links between obesity and atherosclerosis. *Am. J. Physiol. Heart Circ. Physiol.* **288**, H2031–H2041 (2005).
- Poirier, P. *et al.* American Heart Association; Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism. Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss: an update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism. *Circulation* **113**, 898–918 (2006).
- Peeters, A. *et al.* Obesity in adulthood and its consequences for life expectancy: a life-table analysis. *Ann. Intern. Med.* **138**, 24–32 (2003).
- Stevens, J., Cai, J., Evenson, K. R. & Thomas, R. Fitness and fatness as predictors of mortality from all causes and from cardiovascular disease in men and women in the lipid research clinics study. *Am. J. Epidemiol.* **156**, 832–841 (2002).
- Pontiroli, A. E. *et al.* Body weight and glucose metabolism have a different effect on circulating levels of ICAM-1, E-selectin, and endothelin-1 in humans. *Eur. J. Endocrinol.* **150**, 195–200 (2004).
- Blann, A. D., Steele, C. & McCollum, C. N. The influence of smoking on soluble adhesion molecules and endothelial cell markers. *Thromb. Res.* **85**, 433–438 (1997).
- Iwashima, Y. *et al.* Association of hypoadiponectinemia with smoking habit in men. *Hypertension* **45**, 1094–1100 (2005).
- Matsuzawa, Y. The metabolic syndrome and adipocytokines. *FEBS Lett.* **580**, 2917–2921 (2006).
- Liu, R., Mizuta, M. & Matsukura, S. The expression and functional role of nicotinic acetylcholine receptors in rat adipocytes. *J. Pharmacol. Exp. Ther.* **310**, 52–58 (2004).
- Faccini, F. S., Hollenbeck, C. B., Jeppesen, J., Chen, Y. D. & Reaven, G. M. Insulin resistance and cigarette smoking. *Lancet* **339**, 1128–1130 (1992).
- van der Vaart, H., Postma, D. S., Timens, W. & Ten Hacken, N. H. T. Acute effects of cigarette smoking on plasma and oxidative stress: a review. *Thorax* **59**, 713–721 (2004).
- Howard, B. V., Ruotolo, G. & Robbins, D. C. Obesity and dyslipidemia. *Endocrinol. Metab. Clin. N. Am.* **32**, 855–867 (2003).

21. Yusuf, S. *et al.* INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* **364**, 937–952 (2004).
22. Gaudet, D. *et al.* Relationships of abdominal obesity and hyperinsulinemia to angiographically assessed coronary artery disease in men with known mutations in the LDL receptor gene. *Circulation* **97**, 871–877 (1998).
23. Nicholls, S. J. *et al.* Effects of obesity on lipid-lowering, anti-inflammatory and antiatherosclerotic benefits of atorvastatin or pravastatin in patients with coronary artery disease (from the REVERSAL study). *Am. J. Cardiol.* **97**, 1553–1557 (2006).
24. Kwtierovich, P. O. Clinical relevance of the biochemical, metabolic, and genetic factors that influence low-density lipoprotein heterogeneity. *Am. J. Cardiol.* **90**, 30i–47i (2002).
25. Taskiran, M. R. Lipoprotein lipase in diabetes. *Diabetes Metab. Rev.* **3**, 551–570 (1987).
26. Howard, B. V. *et al.* LDL cholesterol as a strong predictor of coronary heart disease in diabetic individuals with insulin resistance and low LDL: The Strong Heart Study. *Arterioscler. Thromb. Vasc. Biol.* **20**, 830–835 (2000).
27. Chan, D. C., Watts, G. F., Redgrave, T. G., Mori, T. A. Barret, P. H. R. Apolipoprotein B-100 kinetics in visceral obesity: associations with plasma apolipoprotein C-III concentration. *Metabolism* **51**, 1041–1046 (2002).
28. Frenais, R. *et al.* *In vivo* evidence for the role of lipoprotein lipase activity in the regulation of apolipoprotein AI metabolism: a kinetic study in control subjects and patients with type II diabetes mellitus. *J. Clin. Endocrinol. Metab.* **86**, 1962–1967 (2001).
29. Mertens, I. & Van Gaal, L. F. New International Diabetes Federation (IDF) and National Cholesterol Education Program Adult Treatment panel III (NCEP-ATPIII) criteria and the involvement of hemostasis and fibrinolysis in the metabolic syndrome. *J. Thromb. Haemost.* **4**, 1164–1166 (2006).
30. Blair, S. N. & Church, T. S. The fitness, obesity, and health equation: is physical activity the common denominator? *J. Am. Med. Assoc.* **292**, 1232–1234 (2004).
31. Li, T. Y. *et al.* Obesity as compared with physical activity in predicting risk of coronary heart disease in women. *Circulation* **113**, 499–506 (2006).
32. Manson, J. E. *et al.* Body weight and mortality among women. *N. Engl. J. Med.* **333**, 677–685 (1995).
33. Sandvik, L. *et al.* Physical fitness as a predictor of mortality among healthy, middle-aged Norwegian men. *N. Engl. J. Med.* **328**, 533–537 (1993).
34. Nigro, J., Osman, N., Dart, A. M. & Little, P. J. Insulin resistance and atherosclerosis. *Endocr. Rev.* **27**, 242–259 (2006).
35. Ferrannini, E. & Izzo, P. Is insulin resistance atherogenic? A review of the evidence. *Atheroscler. Suppl.* **7**, 5–10 (2006).
36. Smith, S. R. & Wilson, P. W. Free fatty acids and atherosclerosis — guilty or innocent? *J. Clin. Endocrinol. Metab.* **91**, 2506–2508 (2006).
37. Pilz, S. *et al.* Free fatty acids are independently associated with all-cause and cardiovascular mortality in subjects with coronary artery disease. *J. Clin. Endocrinol. Metab.* **91**, 2542–2547 (2006).
38. Graham, T. E. *et al.* Retinol-binding protein 4 and insulin resistance in lean, obese, and diabetic subjects. *N. Engl. J. Med.* **354**, 2552–2563 (2006).
39. Abel, E. D. *et al.* Adipose-selective targeting of the *GLUT4* gene impairs insulin action in muscle and liver. *Nature* **409**, 729–733 (2001).
40. Summers, S. A. *et al.* Ceramides in insulin resistance and lipotoxicity. *Prog. Lipid Res.* **45**, 42–72 (2006).
41. Rexrode, K. M. *et al.* Abdominal adiposity and coronary heart disease in women. *J. Am. Med. Assoc.* **280**, 1843–1848 (1998).
42. Després, J. P., Lemieux, I. & Prud'homme, D. Treatment of obesity: need to focus on high risk abdominally obese patients. *Br. Med. J.* **322**, 716–720 (2001).
43. Kuk, J. L. *et al.* Visceral fat is an independent predictor of all-cause mortality in men. *Obesity* **14**, 336–342 (2006).
44. Weisberg, S. P. *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *J. Clin. Invest.* **112**, 1796–1808 (2003).
45. Cencello, R. & Clement, K. Is obesity an inflammatory illness? Role of low-grade inflammation and macrophage infiltration in human white adipose tissue. *BJOG* **113**, 1141–1147 (2006).
46. Bornstein, S. R. *et al.* Immunohistochemical and ultrastructural localization of leptin and leptin receptor in human white adipose tissue and differentiating human adipose cells in primary culture. *Diabetes* **49**, 532–538 (2000).
47. Björntorp, P. 'Portal' adipose tissue as a generator of risk factors for cardiovascular disease and diabetes. *Arteriosclerosis* **10**, 493–496 (1990).
48. Yki-Jarvinen, H. & Westerbacka, J. The fatty liver and insulin resistance. *Curr. Mol. Med.* **5**, 287–295 (2005).
49. Yusuf, S. *et al.* INTERHEART Study Investigators. Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. *Lancet* **366**, 1640–1649 (2005).
50. Montani, J.-P. *et al.* Ectopic fat storage in heart, blood vessels and kidneys in the pathogenesis of cardiovascular diseases. *Int. J. Obesity* **28**, S58–S65 (2004).
51. Zhou, Y. T. *et al.* Lipotoxic heart disease in obese rats: implications for human obesity. *Proc. Natl Acad. Sci. USA* **97**, 1784–1789 (2000).
52. McGavock, J. M., Victor, R. G., Unger, R. H. & Szczepaniak, L. S. American College of Physicians and the American Physiological Society. Adiposity of the heart, revisited. *Ann. Intern. Med.* **144**, 517–524 (2006).
53. Szczepaniak, L. S. *et al.* Myocardial triglycerides and systolic function in humans: *in vivo* evaluation by localized proton spectroscopy and cardiac imaging. *Magn. Reson. Med.* **49**, 417–423 (2003).
54. Verloren, S. *et al.* Visceral periaortic adipose tissue regulates arterial tone of mesenteric arteries. *Hypertension* **44**, 271–276 (2004).
55. Arita, Y. *et al.* Adipocyte-derived plasma protein adiponectin acts as a platelet-derived growth factor-BB-binding protein and regulates growth factor-induced common postreceptor signal in vascular smooth muscle cell. *Circulation* **105**, 2893–2898 (2002).
56. Konstantinides, S., Schafer, K., Koschnick, S. & Loskutoff, D. J. Leptin-dependent platelet aggregation and arterial thrombosis suggests a mechanism for atherothrombotic disease in obesity. *J. Clin. Invest.* **108**, 1533–1540 (2001).
57. Singhal, A. *et al.* Influence of leptin on arterial distensibility: a novel link between obesity and cardiovascular disease? *Circulation* **106**, 1919–1924 (2002).
58. Mertens, I., Considine, R. V., Van der Planken, M. & Van Gaal, L. F. Hemostasis and fibrinolysis in non-diabetic overweight and obese men and women. Is there still a role for leptin? *Eur. J. Endocrinol.* **155**, 477–484 (2006).
59. Kumada, M. *et al.* Adiponectin specifically increases tissue inhibitor of metalloproteinase-1 through interleukin-10 expression in human macrophages. *Circulation* **109**, 2046–2049 (2004).
60. Kobayashi, H. *et al.* Selective suppression of endothelial cell apoptosis by the high molecular weight form of adiponectin. *Circ. Res.* **94**, e27–e31 (2004).
61. Ouchi, N. *et al.* Association of hypoadiponectinemia with impaired vasoreactivity. *Hypertension* **42**, 231–232 (2003).
62. Fukuhara, A. *et al.* Visfatin: a protein secreted by visceral fat that mimics the effects of insulin. *Science* **307**, 426–430 (2005).
63. Haider, D. G. *et al.* The release of the adipocytokine visfatin is regulated by glucose and insulin. *Diabetologia* **49**, 1909–1914 (2006).
64. Mertens, I. & Van Gaal, L. F. Obesity, haemostasis and the fibrinolytic system. *Obes. Rev.* **3**, 85–101 (2002).
65. Bastelica, D. *et al.* Stromal cells are the main plasminogen activator inhibitor-1-producing cells in human fat: evidence of differences between visceral and subcutaneous deposits. *Arterioscler. Thromb. Vasc. Biol.* **22**, 173–178 (2002).
66. Alessi, M. C. *et al.* Plasma PAI-1 levels are more strongly related to liver steatosis than to adipose tissue accumulation. *Arterioscler. Thromb. Vasc. Biol.* **23**, 1262–1268 (2003).
67. Alessi, M. C. & Juhan-Vague, I. PAI-1 and the metabolic syndrome: links, causes, and consequences. *Arterioscler. Thromb. Vasc. Biol.* **26**, 200–207 (2006).
68. Esmon, C. T. The interactions between inflammation and coagulation. *Br. J. Haematol.* **131**, 417–430 (2005).
69. Hanley, A. J. *et al.* Metabolic and inflammation variable clusters and prediction of type 2 diabetes: factor analysis using directly measured insulin sensitivity. *Diabetes* **53**, 1773–1781 (2004).
70. Devaraj, S., Xu, D. Y. & Jialal, I. C-reactive protein increases plasminogen activator inhibitor-1 expression and activity in human aortic endothelial cells: implications for the metabolic syndrome and atherothrombosis. *Circulation* **107**, 398–404 (2003).
71. Ceriello, A. & Motz, E. Is oxidative stress the pathogenic mechanism underlying insulin resistance, diabetes, and cardiovascular disease? The common soil hypothesis revisited. *Arterioscler. Thromb. Vasc. Biol.* **24**, 816–823 (2004).
72. Holvoet, P. *et al.* The metabolic syndrome, circulating oxidized LDL, and risk of myocardial infarction in well-functioning elderly people in the health, aging and body composition cohort. *Diabetes* **53**, 1068–1073 (2004).
73. Maddux, B. A. *et al.* Protection against oxidative stress-induced insulin resistance in rat L6 muscle cells by micromolar concentrations of alpha-lipoic acid. *Diabetes* **50**, 404–410 (2001).
74. Esposito, K. *et al.* Inflammatory cytokine concentrations are acutely increased by hyperglycemia in humans: role of oxidative stress. *Circulation* **106**, 2067–2072 (2002).
75. Tripathy, D. *et al.* Elevation of free fatty acids induces inflammation and impairs vascular reactivity in healthy subjects. *Diabetes* **52**, 2882–2887 (2003).
76. Sattar, N. *et al.* Elevated alanine aminotransferase predicts new-onset type 2 diabetes independently of classical risk factors, metabolic syndrome, and C-reactive protein in the West of Scotland Coronary Prevention Study. *Diabetes* **53**, 2855–2860 (2004).
77. Libby, P., Ridker, P. & Maseri, A. Inflammation and atherosclerosis. *Circulation* **105**, 1135–1143 (2002).
78. Berg, A. H., Scherer, P. E. Adipose tissue, inflammation, and cardiovascular disease. *Circ. Res.* **96**, 939–949 (2005).
79. Berg, A. H., Lin, Y., Lisanti, M. P. & Scherer, P. E. Adipocyte differentiation induces dynamic changes in NF- κ B expression and activity. *Am. J. Physiol. Endocrinol. Metab.* **287**, E1178–E1188 (2004).
80. Rutter, M. K. *et al.* Impact of glucose intolerance and insulin resistance on cardiac structure and function: sex-related differences in the Framingham Heart Study. *Circulation* **107**, 448–454 (2003).
81. Young, M. E. *et al.* Adaptation and maladaptation of the heart in diabetes: part II: potential mechanisms. *Circulation* **105**, 1861–1870 (2002).

Acknowledgement The authors thank A. Verrijken and P.-J. Van Gaal from the Metabolic Unit for their help with literature searches.

Author information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to L.F.V.G. (luc.van.gaal@uza.be).

Abdominal obesity and metabolic syndrome

Jean-Pierre Després^{1,2} & Isabelle Lemieux¹

Metabolic syndrome is associated with abdominal obesity, blood lipid disorders, inflammation, insulin resistance or full-blown diabetes, and increased risk of developing cardiovascular disease. Proposed criteria for identifying patients with metabolic syndrome have contributed greatly to preventive medicine, but the value of metabolic syndrome as a scientific concept remains controversial. The presence of metabolic syndrome alone cannot predict global cardiovascular disease risk. But abdominal obesity — the most prevalent manifestation of metabolic syndrome — is a marker of 'dysfunctional adipose tissue', and is of central importance in clinical diagnosis. Better risk assessment algorithms are needed to quantify diabetes and cardiovascular disease risk on a global scale.

Metabolic syndrome is associated with an increased risk of type 2 diabetes and cardiovascular disease (CVD)^{1–8}. Although there is a debate surrounding the concept of metabolic syndrome^{9–11}, it is recognized as a major and prevalent CVD risk factor by bodies such as the World Health Organization (WHO)¹², the National Cholesterol Education Program–Adult Treatment Panel III (NCEP–ATP III)¹³ and the International Diabetes Federation (IDF)¹⁴.

After an initial attempt by the WHO to define metabolic syndrome on the basis of certain criteria that included an insulin resistance marker, the NCEP–ATP III proposed simple screening tools and cut-off values to help identify patients who are likely to have features of metabolic syndrome and be at increased relative risk of type 2 diabetes and CVD¹³. The five screening variables used to identify those with metabolic syndrome are waist circumference, circulating levels of triacylglycerols and of high-density lipoprotein (HDL)-cholesterol, fasting glycaemia and blood pressure. A meta-analysis of the prospective studies that have used these criteria has shown that the presence of metabolic syndrome increases the risk of type 2 diabetes and CVD¹⁵.

The recommendation to measure waist circumference rather than body mass index (BMI) recognized the important part played by abdominal obesity in metabolic syndrome. By singling out waist circumference, the NCEP–ATP III recommendation acknowledged that health professionals and clinicians are struggling with a demographic explosion: more and more patients are overweight or obese and show the related metabolic effects of an affluent, sedentary lifestyle characterized by excess consumption of highly processed, energy-dense food of poor nutritional value. The parallel rapid growth of overweight and obese individuals and of type 2 diabetes is striking¹⁶, and led to the coining of the term 'diabesity'^{17–20} to emphasize the link between these two conditions. Type 2 diabetes might be a significant CVD risk factor, but the independent contribution of the hyperglycaemic state of type 2 diabetes to CVD risk is rather weak. This hyperglycaemic state is only the tip of a huge dysmetabolic iceberg, mostly resulting from a combination of factors found in overweight and obese patients with excess abdominal fat and insulin resistance².

The existence of metabolic syndrome implies a shift from a pathophysiological concept based on metabolic abnormalities resulting from an insulin-resistant state to an epidemiological construct based on abdominal obesity and crude correlates of the features of insulin resistance. Some investigators argue that there is no rationale for the

existence of metabolic syndrome and that insulin resistance is the key culprit increasing CVD risk (on top of traditional risk factors) in obese individuals^{7,9–11}. Non-Caucasian populations may be more or less prone to abdominal and visceral fat accumulation^{21–26} and to the development of metabolic abnormalities, and it has been proposed that the publication of specific waist circumference cut-offs to define abdominal obesity in various ethnic groups is not supported by solid epidemiological and metabolic data.

Beyond excess body weight

Although obesity is a risk factor for insulin resistance and type 2 diabetes, and a significant risk factor for CVD, not every obese patient is insulin resistant²⁷ or at high risk of diabetes and CVD. This explains why obesity has been an ill-defined modifiable CVD risk factor compared with others such as hypertension, smoking and cholesterol (high low-density lipoprotein (LDL)/low HDL). But, for any given amount of total body fat, the subgroup of individuals with a selective excess of intra-abdominal, or visceral, adipose tissue is at substantially higher risk of being characterized by insulin resistance and by the features of metabolic syndrome^{28,29}. Although excess visceral fat accumulation is associated with various atherogenic and diabetogenic abnormalities^{28–32}, an important question has been whether visceral fat is a causal factor or simply a marker of a dysmetabolic profile.

Pathophysiology of visceral obesity

There is ample evidence that an impaired non-esterified fatty acid (NEFA) metabolism could contribute to the insulin-resistant state observed among individuals with visceral obesity. Hypertrophied intra-abdominal adipocytes are characterized by a hyperlipolytic state that is resistant to the antilipolytic effect of insulin^{33,34}. The resulting NEFA flux to the liver may impair liver metabolism, leading to increased hepatic glucose production. Hepatic insulin resistance is associated with decreased apolipoprotein B degradation and increased production of triacylglycerol-rich lipoproteins. A high-fat diet promoting visceral adiposity in a canine model of diet-induced visceral obesity can induce hepatic insulin resistance with respect to glucose production, whereas the sensitivity of peripheral tissues seems to be less affected by the diet-induced increase in visceral adiposity³⁵. Although trivial differences in fasting NEFA levels have sometimes been observed in response to this high-fat regimen, there was a marked increase in the 24-hour NEFA

¹Québec Heart Institute, Hôpital Laval Research Centre, 2725 chemin Sainte-Foy, Pavillon Marguerite-D'Youville, 4th Floor, Quebec City, QC G1V 4G5, Canada. ²Division of Kinesiology, Department of Social and Preventive Medicine, Université Laval, Quebec City, QC G1K 7P4, Canada

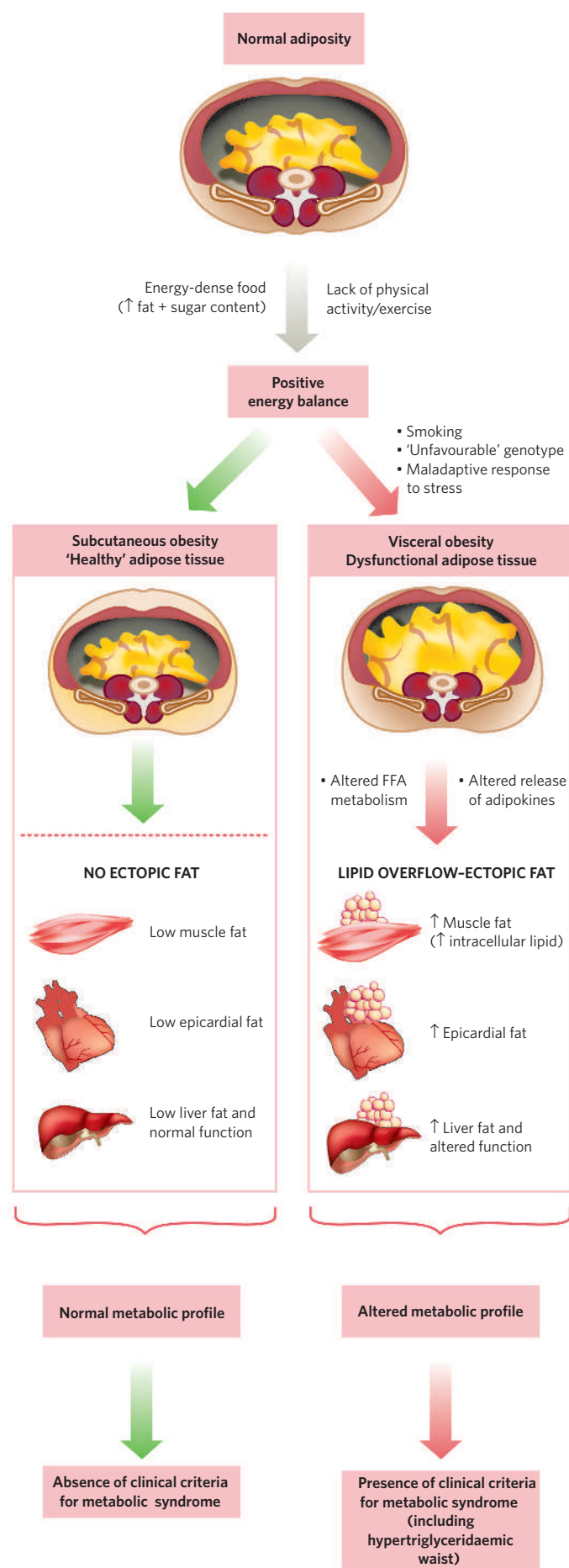


Figure 1 | The lipid overflow-ectopic fat model. Excess visceral fat accumulation might be causally related to the features of insulin resistance, but might also be a marker of a dysfunctional adipose tissue being unable to appropriately store the energy excess. According to this model, the body's ability to cope with the surplus of calories (resulting from excess caloric consumption, a sedentary lifestyle or, as is often the case, a combination of both factors) might, ultimately, determine the individual's susceptibility to developing metabolic syndrome. There is evidence suggesting that if the extra energy is channelled into insulin-sensitive subcutaneous adipose tissue, the individual, although in positive energy balance, will be protected against the development of the metabolic syndrome. However, in cases in which adipose tissue is absent, deficient or insulin resistant with a limited ability to store the energy excess, the triacylglycerol surplus will be deposited at undesirable sites such as the liver, the heart, the skeletal muscle and in visceral adipose tissue — a phenomenon described as ectopic fat deposition. Factors associated with a preferential accumulation of visceral fat and with features of insulin resistance include, among others, smoking, the well-documented genetic susceptibility to visceral obesity³⁴ and a neuroendocrine profile related to a maladaptive response to stress³⁰. The resulting metabolic consequences of this 'defect' in energy partitioning include visceral obesity, insulin resistance, an atherogenic dyslipidaemia and a pro-thrombotic, inflammatory profile. These are defining features of metabolic syndrome. This constellation of abnormalities can be detected by the clinical criteria for metabolic syndrome, the two simplest being the simultaneous presence of increased waist girth and fasting triacylglycerol levels, a condition that has been described as 'hypertriglyceridaemic waist'⁸⁵.

profile of these visceraally obese dogs. It was therefore proposed³⁵ that such an increase in NEFAs could be a stimulus for insulin secretion and could have an important role in the aetiology of insulin resistance, particularly as it relates to hepatic carbohydrate and lipid metabolism.

In humans, although there is a correlation between visceral fat accumulation and portal delivery of NEFAs to the liver, most portal NEFAs originate from the systemic circulation. This suggests that other factors might explain the altered metabolic profile of visceraally obese patients³⁶. There is evidence that adipose tissue is not only specialized in the storage and mobilization of lipids but that it is also a remarkable endocrine organ releasing numerous cytokines, including, among many others, proinflammatory molecules such as interleukin (IL)-6 and tumour necrosis factor- α (TNF- α). In obesity, there is evidence of macrophage infiltration in adipose tissue³⁷, which could contribute to the inflammatory profile that has been reported in abdominally obese patients³⁸. Plasma levels of C-reactive protein (CRP), an inflammatory marker that is predictive of a risk of myocardial infarction possibly greater than that estimated by traditional risk factors³⁹, are increased in patients with visceral obesity⁴⁰.

The protein adiponectin^{41,42} is abundant in the blood, and is specifically derived from adipose tissue. As opposed to proinflammatory adipokines, adiponectin levels are reduced in obese individuals, particularly among patients with excess visceral adiposity⁴³. Adiponectin has been found to have many effects *in vitro* that are compatible with improved insulin signalling and potential protection against atherosclerosis^{44,45}. The reduced adiponectin levels observed in visceraally obese patients could therefore be one of the key factors responsible for their atherogenic and diabetogenic metabolic risk factor profile. Abdominally obese patients with an excess of visceral adipose tissue have elevated plasma CRP concentrations accompanied by elevated IL-6 and TNF- α levels and by reduced adiponectin concentrations^{43,46}. However, although low adiponectin levels are a salient feature of visceral obesity, whether this adipokine has a central role in the altered metabolic risk profile of patients with visceral obesity remains uncertain. Overall, these results are consistent with an important endocrine function of the expanded visceral adipose depot not only leading to altered NEFA metabolism but also to a proinflammatory profile⁴⁷ that might contribute to the insulin resistance and altered glucose homeostasis of visceraally obese patients⁴⁸.

Both the altered NEFA metabolism and the endocrine function hypotheses imply that visceral adipose tissue is causally involved in the pathophysiology of the metabolic syndrome that is often found in

patients with visceral obesity. However, another possibility (which does not exclude a contribution from the two mechanisms described above) is that excess intra-abdominal fat accumulation represents a marker of the relative inability of subcutaneous adipose tissue to act as an 'energy sink' when an individual has to handle a calorie surplus due to excess energy intake and/or reduced energy expenditure⁴⁹ (Fig. 1). Such a relative deficit in the capacity of subcutaneous fat to store excess energy would result in increased accumulation of fat at undesired sites such as the liver, the skeletal muscle, the heart and even in pancreatic β -cells, a phenomenon that has been described as ectopic fat deposition⁵⁰. Consistent with this theory is the fact that transgenic mice that are essentially fatless owing to the expression of A-ZIP/F-1 protein — which blocks the activity of several transcription factors — also show liver and muscle insulin resistance and eventually develop diabetes^{51,52}. Surgical implantation of adipose tissue in these mice improves the insulin sensitivity of their liver and muscles^{51,52}, consistent with the idea that subcutaneous fat is a metabolic sink to buffer an energy surplus. In humans, the severe insulin-resistant state found in patients with lipodystrophic conditions⁵³ is also consistent with the role of subcutaneous adipose tissue as a depot buffering the energy excess⁵⁴. In accordance with this hypothesis, treatment with glitazones increases subcutaneous fat deposition⁵⁵, which might help to explain the beneficial effects of this class of drug on muscle and liver insulin sensitivity.

Thus, the insulin-resistant, dyslipidaemic state found in patients with the features of metabolic syndrome might be only partly explained by the peculiar metabolic and endocrine properties of the expanded visceral adipose tissue. Visceral obesity might also be a marker of defective fat partitioning between the adipose tissue, the skeletal muscle, the liver and the heart.

On the basis of the association between abdominal, especially visceral, adiposity and the presence of the features of metabolic syndrome, the measurement of waist circumference has been proposed as a crude anthropometric correlate of abdominal and visceral adiposity^{12–14}. But measuring waist girth has its limitations, which are discussed below.

Clinical criteria

The five simple criteria and cut-off values proposed by the NCEP-ATP III panel^{2,13} and endorsed by the IDF¹⁴ to diagnose the likely presence of metabolic syndrome were reached through expert consensus and interpretation of the literature. The criteria are: increased waist circumference

with population-specific cut-off values; increased triacylglycerol levels or treatment for hypertriglyceridaemia; low HDL-cholesterol concentration or treatment for this condition; elevated blood pressure or treatment for hypertension; and elevated glucose concentration or treatment with a hypoglycaemic agent. These criteria and values have not been validated for their ability to discriminate optimally for individuals with both metabolic syndrome and a related increase in CVD risk. This is particularly crucial for the assessment of CVD risk associated with excess visceral adiposity in non-Caucasian populations, an area in which much work is needed. Despite this limitation, prospective observational studies have generally shown that individuals who meet the clinical criteria for metabolic syndrome are at increased risk of CVD events and type 2 diabetes compared with individuals without the syndrome¹⁵. Using different cut-offs or metabolic syndrome markers might improve identification of patients at increased risk. For instance, an elevated fasting blood glucose concentration, which is often referred to as a 'prediabetic' state, is more useful for predicting type 2 diabetes risk^{56–60} than CVD risk^{61–63}.

Since their publication, the conceptual definition of metabolic syndrome has often been confused with the five proposed clinical criteria. These criteria are merely surrogate variables to help identify a subgroup of high-risk individuals likely to be characterized by key features of the metabolic syndrome: abdominal obesity, insulin resistance, high triacylglycerol-apolipoprotein B, low HDL-cholesterol, small, dense LDL dyslipidaemia, a pro-thrombotic state and an inflammatory profile. These clustering features are the most prevalent form of the metabolic syndrome as defined by NCEP-ATP III⁶⁴. This constellation of abnormalities might be accompanied by hypertension and/or type 2 diabetes, depending on the individual's genetic susceptibility.

Despite some limited evidence^{56,65}, it remains uncertain whether all possible combinations of three of the five NCEP-ATP III criteria similarly increase CVD risk. This warrants further attention.

Limitations

One limitation of metabolic syndrome is that although it leads to an approximately twofold increase in relative CVD risk¹⁵, it should not replace the need to assess overall cardiovascular risk taking into account well-established CVD risk factors such as age, gender, smoking, blood pressure, cholesterol (or LDL-cholesterol) and diabetes¹⁰. It has also

Box 1 | Assessment of global CHD risk

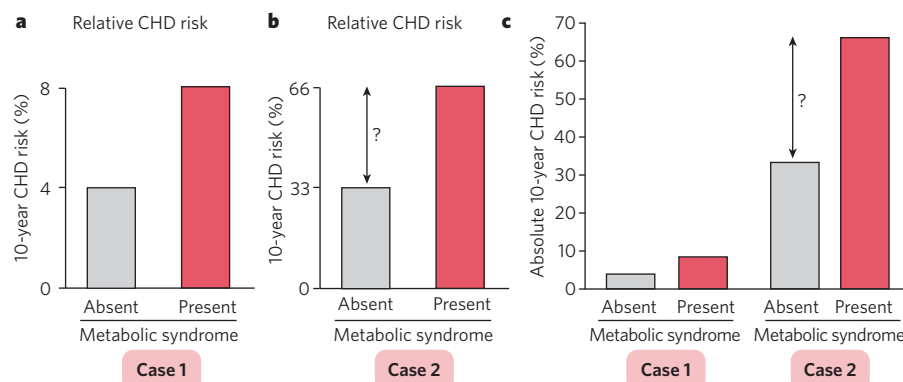
Meeting the clinical criteria for the metabolic syndrome does not necessarily equal a very high absolute risk of CVD. Consider the following two cases.

The first case (panel **a**) is a 30-year-old nondiabetic man with normal blood pressure, who has never smoked and has an LDL-cholesterol concentration of 3.00 mmol l⁻¹. The patient has a waist girth of 104 cm, fasting plasma triacylglycerol levels of 2.50 mmol l⁻¹, and an HDL-cholesterol concentration of 0.85 mmol l⁻¹. Although this patient is diagnosed as having metabolic syndrome (its presence presumably doubling his risk of CHD), his Framingham score — which estimates 10-year CHD risk on the basis of traditional risk factors — is low (4%). This is because the patient is young and does not have the traditional risk factors, with the exception of a low HDL-cholesterol concentration, a prevalent correlate of abdominal obesity. A twofold increase in risk due to the presence of metabolic syndrome therefore produces an absolute CHD risk of only 8%.

Let us now consider the second case (panel **b**), who meets exactly the same NCEP-ATP III

metabolic syndrome criteria as the first (nondiabetic, waist girth of 104 cm, fasting plasma triacylglycerol levels of 2.50 mmol l⁻¹ and an HDL-cholesterol concentration of 0.85 mmol l⁻¹). However, this patient is 55 years of age, smokes one pack of cigarettes per day, has untreated hypertension (170/95 mmHg) and has an LDL-cholesterol concentration of 4.10 mmol l⁻¹. Under the

Framingham model, this patient is at much greater absolute risk of CHD than the patient in case number 1 (33% compared with 4%), irrespective of metabolic syndrome. Thus, the same theoretical twofold increase in CHD risk due to metabolic syndrome would generate a much higher absolute CHD risk in the second patient (66%) than the first (8%). This is illustrated by panel **c**.



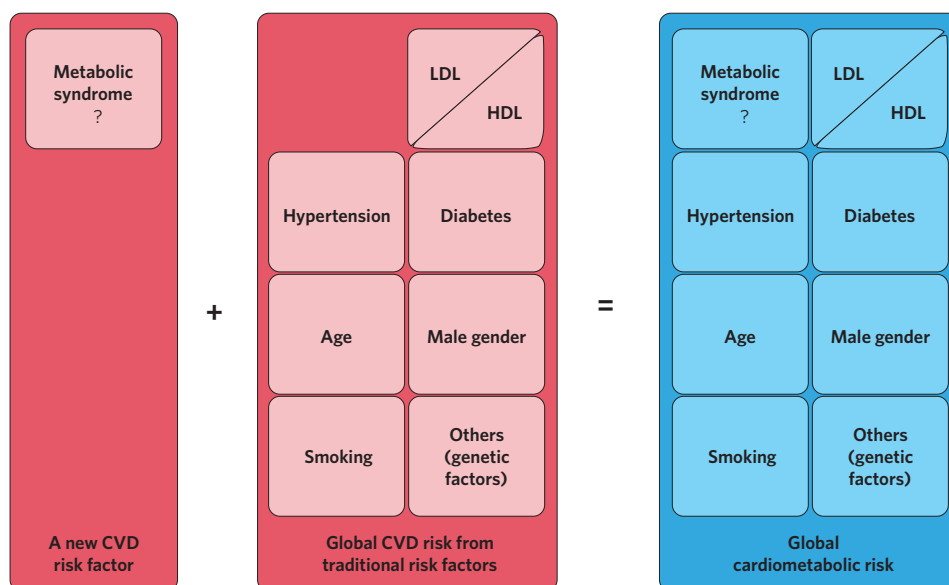


Figure 2 | Factors contributing to global cardiometabolic risk. Cardiometabolic risk is the overall risk of CVD resulting from the presence of metabolic syndrome but also of traditional risk factors such as lipids (LDL and HDL), hypertension, diabetes, age, male gender, smoking and other unknown risk factors (including genetic factors that cannot be assessed in clinical practice most of the time). According to this model, metabolic syndrome does not replace the need to assess global CVD risk, but may eventually have to be considered in global risk assessment. Whether metabolic syndrome is an independent factor that adds significantly to the global CVD risk as assessed with traditional risk factors is uncertain and much debated in the literature. The controversy over its added value is highlighted by the question mark.

been argued that current risk-assessment algorithms such as the Framingham Heart Study calculator of coronary heart disease (CHD) risk⁶⁶ largely capture the risk associated with metabolic syndrome^{67,68}. The Framingham Heart Study calculator, for instance, considers some elements of metabolic syndrome, such as blood pressure and HDL-cholesterol levels. Although the presence of clinical criteria for metabolic syndrome is predictive of an increased relative CVD risk, the absolute risk of CVD is mainly determined by the presence or absence of traditional risk factors. Therefore, meeting the clinical criteria for metabolic syndrome does not necessarily equal a very high absolute risk of CVD. The two cases presented in Box 1 illustrate this point.

When assessing the absolute CVD risk of patients who meet the clinical criteria for metabolic syndrome, it is necessary first to pay attention to traditional risk factors. This is one of the key criticisms of metabolic syndrome and its relevance to clinical practice^{9,10}. A lively debate is currently ongoing as to whether metabolic syndrome enhances our understanding of global CHD risk as assessed by available algorithms such as the Framingham⁶⁶ or PROCAM⁶⁹ (Prospective Cardiovascular Münster study) risk calculators. Conflicting and mostly negative results have been reported^{67,68,70}, and further studies are needed. But some evidence suggests that when sophisticated markers of metabolic syndrome — such as fasting insulin and apolipoprotein B levels, and LDL size — are measured, their presence increases CVD risk beyond that which would be calculated using traditional CVD risk algorithms⁷¹. For the time being, it is important to emphasize that the diagnosis of metabolic syndrome does not necessarily entail a high absolute risk of CVD. Rather, such a diagnosis should be a cause for concern, as it identifies an individual with a dysfunctional metabolism who needs to change his or her lifestyle and lose weight, especially abdominal fat. Thus, diagnosis of metabolic syndrome does not automatically identify a candidate for pharmacotherapy, nor should it detract from the importance of pharmacological management of traditional risk factors in accordance with current guidelines^{1,9}.

Evolving guidelines

The IDF working group proposed metabolic syndrome criteria that conformed with NCEP–ATP III recommendations^{13,14}. The IDF committee placed special emphasis on abdominal obesity and the measurement of waist circumference. Their guidelines proposed that increased waist circumference was a necessary criterion to identify patients at risk of having metabolic syndrome. Furthermore, in light of compelling evidence that the waist girth cut-off value proposed for men by NCEP–ATP III (102 cm) was too high, particularly in some ethnic populations, the IDF guidelines reduced critical waist values to 94 cm

in men and 80 cm in women, noting that factors such as ethnicity and age substantially affect the relationship of waist to abdominal visceral fat deposition and related abnormalities. Given that most patients with metabolic syndrome have an excess of abdominal fat, it was a step forward when increased waist girth was included in the IDF guidelines as the first criterion to identify individuals likely to have metabolic syndrome. But the focus on abdominal obesity has been criticized by some who feel that insulin resistance is the syndrome's central component^{7,10}, given that there are forms of insulin resistance not related to excess abdominal fat^{53,54,72}.

This debate has caused confusion in the media and the medical field, and it is unfortunate that major organizations have been unable to reach a consensus to emphasize that the most prevalent form of insulin resistance is found among patients with an excess of abdominal fat²⁹. Of course, smoking is a confounding factor in the abdominal obesity–insulin resistance association, as smokers have more abdominal adipose tissue than non-smokers even after adjustments for markers of total adiposity^{73,74}. Smoking is also associated with insulin resistance and low HDL-cholesterol levels⁷⁵. However, metabolic studies conducted exclusively on non-smokers have confirmed the very significant relationship between abdominal adiposity and an atherogenic and diabetogenic metabolic risk profile^{28,29,76,77}. This supports the idea that smoking is not the main factor responsible for the elevated CVD risk of abdominally obese individuals.

Challenging waist

Another criticism of metabolic syndrome has been the relevance or appropriateness of lowering waist circumference cut-off values used to define abdominal obesity in the recent IDF guidelines¹⁴. In some parts of the world, this reduction in waist circumference cut-offs has considerably increased the number of patients being diagnosed with metabolic syndrome. Some have even seen this as an attempt to increase the number of patients who would be eligible for pharmacotherapy^{1,9}. To close this debate, clinical endpoints (incidence of diabetes and CVD) and prospective international epidemiological data with proper body-fat distribution and metabolic markers are urgently needed to quantify the risk associated with features of metabolic syndrome in various populations of the world. For instance, a study conducted in Greenland Inuit had indicated that this population might be less prone to developing metabolic syndrome in the presence of obesity than Caucasian whites⁷⁸. However, no imaging data of visceral adiposity were obtained in this study. In addition, the available evidence suggests that the definition of obesity is markedly different in Asia compared with North America or Europe²³.

Some ethnic populations are more or less susceptible to visceral fat accumulation for a given amount of total body fat^{21–26}. The definition of high-risk abdominal obesity must, therefore, be tailored to various world populations. This issue is a serious roadblock to properly quantifying the prevalence of metabolic syndrome and its clinical sequelae. To this end, modern metabolic epidemiology studies are needed with comprehensive metabolic measurements and proper assessment of regional body fatness using imaging techniques. These studies should quantify abdominal subcutaneous and visceral adipose tissue and ectopic fat deposition, as well as assess relevant metabolic markers so that the relationship of metabolic parameters and clinical events to the size of these regional fat deposits can be more accurately measured.

For now, the numerous papers published on metabolic syndrome's estimated worldwide prevalence must be interpreted with great caution as they may be misleading. For this reason, the IDF committee has identified knowledge gaps in their recommendations¹⁴. On this basis, the recently reported 'new' IDF waist-girth criteria should be considered a work in progress, and will have to be validated for their ability to discriminate optimally for the subgroup of individuals who have features of metabolic syndrome and are at increased risk of CVD.

The current epidemic of type 2 diabetes and metabolic syndrome is a direct result of our energy-dense diet and affluent sedentary lifestyle. Because such a lifestyle increases the likelihood of individuals eating more than they need, this positive energy balance leads to abdominal obesity and insulin resistance in the presence of an unfavourable genotype and other permissive factors (such as smoking and a maladaptive response to stress). Failure to consider high-risk abdominal obesity as the most prevalent form of metabolic syndrome will unfortunately confuse many physicians and their patients. Furthermore, it would detract from the attention that should be given to weight loss, which has been reported to induce a substantial mobilization of abdominal and visceral adipose tissue among high-risk abdominally obese patients^{77,79}. Such selective mobilization of abdominal/visceral fat has been suggested to be an important factor in explaining why moderate weight loss improves the metabolic profile of most patients with metabolic syndrome^{79–83}.

Strengths and weaknesses

Not all patients with an increased waist girth have the features of metabolic syndrome, leading some to question its usefulness⁷. Although the finding is not surprising, it has supported the assertion that abdominal obesity is not a component of the syndrome. However, an increased waistline may be the consequence of excess subcutaneous abdominal adiposity, a situation sometimes observed even in very obese patients with a normal risk-factor profile⁸⁴. Once waist circumference has been assessed as a first step, the presence or absence of hypertriglyceridaemia might help to distinguish high-risk visceral obesity from lower-risk subcutaneous obesity (Fig. 1). An increased waist circumference alone is therefore not sufficient to identify a high-risk abdominally obese patient with excess visceral adipose tissue. Clinical markers of an altered metabolic risk profile, such as clinical criteria for metabolic syndrome (the simplest being increased triacylglycerol levels), must also be present to suggest the presence of high-risk visceral obesity^{85–87}. Once classical and metabolic risk factors have been taken into account, there is little evidence that waist circumference alone is an independent risk factor for CVD.

If we consider the pathophysiology of the most prevalent form of metabolic syndrome, then the most prevalent correlate of the metabolic syndrome epidemic is abdominal obesity. Thus, after accounting for the consequences of abdominal obesity (such as insulin resistance, atherogenic dyslipidaemia, hypertension, hyperglycaemia, a pro-thrombotic state and an inflammatory profile), waist circumference is unlikely to predict CVD events. But should we stop assessing abdominal obesity and put the worldwide obesity epidemic on the back burner? The answer to this question is an emphatic 'no'. In the presence of the clinical criteria of metabolic syndrome, an increased waist circumference does provide relevant pathophysiological information insofar as it defines the prevalent form of the syndrome resulting from abdominal obesity.

But a key issue remains unsolved: the identification of high-risk abdominally obese patients in various populations of the world. Even for the US population, there was no scientific rationale behind the waist circumference cut-offs of 102 cm in men and 88 cm in women. These waist girth values corresponded to a BMI of 30 kg m⁻² in men and women, and were simply taken from an earlier European study⁸⁸. Other cut-offs have been proposed on the basis of metabolic markers and disease states^{57,85,87,89}. As mentioned above, the relationship of total adiposity to visceral fat deposition and to metabolic complications may vary between populations. The available data suggest that blacks are more prone to subcutaneous fat accumulation for a given BMI than are whites^{21,22,25,26}, whereas Asians are quite prone to visceral fat accumulation^{23,24}, which may explain their greater propensity to develop diabetes at relatively low BMI values. We therefore need to establish the relationship of anthropometry to visceral and subcutaneous adiposity in various populations, and to study a comprehensive set of metabolic syndrome markers in order to quantify their relationship to specific clinical events such as type 2 diabetes and CVD.

The future of global cardiovascular disease risk assessment

Better global risk-assessment algorithms are needed to quantify diabetes and CVD risk resulting from the presence of classical risk factors and the presence of abdominal obesity or insulin resistance-related metabolic markers. The term 'cardiometabolic risk' has been coined by organizations such as the American Diabetes Association⁹⁰ and the American Heart Association⁹¹ to describe the overall risk of developing type 2 diabetes and CVD^{92,93}, and this idea may potentially reconcile both supporters and detractors of the metabolic syndrome concept. As illustrated in Fig. 2, cardiometabolic risk encompasses the global risk of CVD and type 2 diabetes associated with traditional risk factors while also taking into consideration the potential additional contribution of abdominal obesity and/or insulin resistance and of related metabolic markers (to be identified) to global CVD risk. Current evidence does not suggest that the presence of clinical criteria for metabolic syndrome adds to global CVD risk. Only additional prospective studies, which will consider the measurement of sophisticated metabolic markers and direct measurements of abdominal visceral and subcutaneous adiposity, have the potential to answer this important question. Once these results become available we should be better positioned to address the key questions of what constitutes a high-risk abdominal obesity phenotype in various regions of the world and what the main determinants of risk in different populations are.

However, a distinction must clearly be made between metabolic syndrome as a concept and the criteria used in clinical practice to identify individuals with features of metabolic syndrome. Although insulin resistance is a key component of a constellation of metabolic abnormalities, which increase the risk of type 2 diabetes and CVD, the most prevalent form of insulin resistance is associated with abdominal obesity and with 'dysfunctional' adipose tissue that cannot properly handle the energy surplus resulting from a sedentary lifestyle combined with excessive calorie consumption^{49,54}. Initial indicators of high-risk abdominal obesity are an increased waist circumference along with raised fasting plasma triacylglycerol concentrations⁸⁵. Although metabolic syndrome increases relative CVD risk, its diagnosis does not necessarily mean that a patient is at very high risk of a cardiovascular event. To properly evaluate cardiovascular risk, physicians must first consider traditional CVD risk factors. Whether the presence of the clinical criteria for the metabolic syndrome increases the risk of CVD beyond that of traditional risk factors is not yet clear. Resolving this is crucial for the optimal assessment of global CVD risk.

1. Grundy, S. M. Metabolic syndrome: connecting and reconciling cardiovascular and diabetes worlds. *J. Am. Coll. Cardiol.* **47**, 1093–1100 (2006).
2. Grundy, S. M. Does a diagnosis of metabolic syndrome have value in clinical practice? *Am. J. Clin. Nutr.* **83**, 1248–1251 (2006).
3. Grundy, S. M. Drug therapy of the metabolic syndrome: minimizing the emerging crisis of polypharmacy. *Nature Rev. Drug Discov.* **5**, 295–309 (2006).
4. Moller, D. E. & Kaufman, K. D. Metabolic syndrome: a clinical and molecular perspective. *Annu. Rev. Med.* **56**, 45–62 (2005).

5. Sattar, N. The metabolic syndrome: should current criteria influence clinical practice? *Curr. Opin. Lipidol.* **17**, 404–411 (2006).
6. Eckel, R. H., Grundy, S. M. & Zimmet, P. Z. The metabolic syndrome. *Lancet* **365**, 1415–1428 (2005).
7. Reaven, G. M. The metabolic syndrome: is this diagnosis necessary? *Am. J. Clin. Nutr.* **83**, 1237–1247 (2006).
8. Ferrannini, E. Is insulin resistance the cause of the metabolic syndrome? *Ann. Med.* **38**, 43–51 (2006).
9. Gale, E. A. The myth of the metabolic syndrome. *Diabetologia* **48**, 1679–1683 (2005).
10. Kahn, R., Buse, J., Ferrannini, E. & Stern, M. The metabolic syndrome: time for a critical appraisal: joint statement from the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care* **28**, 2289–2304 (2005).
11. Reaven, G. M. The metabolic syndrome: requiescat in pace. *Clin. Chem.* **51**, 931–938 (2005).
12. Alberti, K. G. & Zimmet, P. Z. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet. Med.* **15**, 539–553 (1998).
13. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). *J. Am. Med. Assoc.* **285**, 2486–2497 (2001).
14. Alberti, K. G., Zimmet, P. & Shaw, J. The metabolic syndrome — a new worldwide definition. *Lancet* **366**, 1059–1062 (2005).
15. Galassi, A., Reynolds, K. & He, J. Metabolic syndrome and risk of cardiovascular disease: a meta-analysis. *Am. J. Med.* **119**, 812–819 (2006).
16. Mokdad, A. H. et al. Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. *J. Am. Med. Assoc.* **289**, 76–79 (2003).
17. Zimmet, P., Alberti, K. G. M. M. & Shaw, J. Global and societal implications of the diabetes epidemic. *Nature* **414**, 782–787 (2001).
18. Astrup, A. & Finer, N. Redefining type 2 diabetes: 'diabesity' or 'obesity dependent diabetes mellitus'? *Obesity Rev.* **1**, 57–59 (2000).
19. Shafir, E. Development and consequences of insulin resistance: lessons from animals with hyperinsulinaemia. *Diabetes Metab.* **22**, 122–131 (1996).
20. Pincock, S. Paul Zimmet: fighting the 'diabesity' pandemic. *Lancet* **368**, 1643 (2006).
21. Després, J. P. et al. Race, visceral adipose tissue, plasma lipids, and lipoprotein lipase activity in men and women: the Health, Risk Factors, Exercise Training, and Genetics (HERITAGE) family study. *Arterioscler. Thromb. Vasc. Biol.* **20**, 1932–1938 (2000).
22. Albu, J. B., Murphy, L., Frager, D. H., Johnson, J. A. & Pi-Sunyer, F. X. Visceral fat and race-dependent health risks in obese nondiabetic premenopausal women. *Diabetes* **46**, 456–462 (1997).
23. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* **363**, 157–163 (2004).
24. Kadowaki, T. et al. Japanese men have larger areas of visceral adipose tissue than Caucasian men in the same levels of waist circumference in a population-based study. *Int. J. Obes. (Lond.)* **30**, 1163–1165 (2006).
25. Lovejoy, J. C., de la Bretonne, J. A., Klemperer, M. & Tulley, R. Abdominal fat distribution and metabolic risk factors: effects of race. *Metabolism* **45**, 1119–1124 (1996).
26. Conway, J. M., Chanetsa, F. F. & Wang, P. Intrabdominal adipose tissue and anthropometric surrogates in African American women with upper- and lower-body obesity. *Am. J. Clin. Nutr.* **66**, 1345–1351 (1997).
27. Abbasi, F., Brown, B. W., Lamendola, C., McLaughlin, T. & Reaven, G. M. Relationship between obesity, insulin resistance, and coronary heart disease risk. *J. Am. Coll. Cardiol.* **40**, 937–943 (2002).
28. Després, J. P. et al. Regional distribution of body fat, plasma lipoproteins, and cardiovascular disease. *Arteriosclerosis* **10**, 497–511 (1990).
29. Després, J. P. Is visceral obesity the cause of the metabolic syndrome? *Ann. Med.* **38**, 52–63 (2006).
30. Björntorp, P. Metabolic implications of body fat distribution. *Diabetes Care* **14**, 1132–1143 (1991).
31. Kissebah, A. H. & Krakower, G. R. Regional adiposity and morbidity. *Physiol. Rev.* **74**, 761–811 (1994).
32. Lebovitz, H. E. & Banerji, M. A. Point: visceral adiposity is causally related to insulin resistance. *Diabetes Care* **28**, 2322–2325 (2005).
33. Mittelman, S. D., Van Citters, G. W., Kirkman, E. L. & Bergman, R. N. Extreme insulin resistance of the central adipose depot in vivo. *Diabetes* **51**, 755–761 (2002).
34. Mauriège, P. et al. Regional variation in adipose tissue metabolism of severely obese premenopausal women. *J. Lipid. Res.* **36**, 672–684 (1995).
35. Bergman, R. N. et al. Why visceral fat is bad: mechanisms of the metabolic syndrome. *Obesity (Silver Spring)* **14** (Suppl. 1), 16S–19S (2006).
36. Jensen, M. D. Is visceral fat involved in the pathogenesis of the metabolic syndrome? Human model. *Obesity (Silver Spring)* **14** (Suppl. 1), 20S–24S (2006).
37. Weisberg, S. P. et al. Obesity is associated with macrophage accumulation in adipose tissue. *J. Clin. Invest.* **112**, 1796–1808 (2003).
38. Yudkin, J. S., Stehouwer, C. D., Emeis, J. J. & Coppack, S. W. C-reactive protein in healthy subjects: associations with obesity, insulin resistance, and endothelial dysfunction: a potential role for cytokines originating from adipose tissue? *Arterioscler. Thromb. Vasc. Biol.* **19**, 972–978 (1999).
39. Tsimikas, S., Willerson, J. T. & Ridker, P. M. C-reactive protein and other emerging blood biomarkers to optimize risk stratification of vulnerable patients. *J. Am. Coll. Cardiol.* **47**, C19–C31 (2006).
40. Lemieux, I. et al. Elevated C-reactive protein: another component of the atherothrombotic profile of abdominal obesity. *Arterioscler. Thromb. Vasc. Biol.* **21**, 961–967 (2001).
41. Maeda, K. et al. cDNA cloning and expression of a novel adipose specific collagen-like factor, apM1 (adipose most abundant gene transcript 1). *Biochem. Biophys. Res. Commun.* **221**, 286–289 (1996).
42. Scherer, P. E., Williams, S., Fogliano, M., Baldini, G. & Lodish, H. F. A novel serum protein similar to C1q, produced exclusively in adipocytes. *J. Biol. Chem.* **270**, 26746–26749 (1995).
43. Côté, M. et al. Adiponectinemia in visceral obesity: impact on glucose tolerance and plasma lipoprotein and lipid levels in men. *J. Clin. Endocrinol. Metab.* **90**, 1434–1439 (2005).
44. Berg, A. H. & Scherer, P. E. Adipose tissue, inflammation, and cardiovascular disease. *Circ. Res.* **96**, 939–949 (2005).
45. Matsuzawa, Y. Therapy Insight: adipocytokines in metabolic syndrome and related cardiovascular disease. *Nature Clin. Pract. Cardiovasc. Med.* **3**, 35–42 (2006).
46. Lemieux, I. et al. Contribution of interleukin-6 and tumor necrosis factor- α to the elevated C-reactive protein levels found in abdominally obese men. *Obes. Res.* **9**, 665 (2001).
47. Hotamisligil, G. S. Molecular mechanisms of insulin resistance and the role of the adipocyte. *Int. J. Obes. Relat. Metab. Disord.* **24** (Suppl. 4), S23–S27 (2000).
48. Havel, P. J. Control of energy homeostasis and insulin action by adipocyte hormones: leptin, acylation stimulating protein, and adiponectin. *Curr. Opin. Lipidol.* **13**, 51–59 (2002).
49. Lemieux, I. Energy partitioning in gluteal-femoral fat: does the metabolic fate of triglycerides affect coronary heart disease risk? *Arterioscler. Thromb. Vasc. Biol.* **24**, 795–797 (2004).
50. Miranda, P. J., DeFronzo, R. A., Califf, R. M. & Guyton, J. R. Metabolic syndrome: definition, pathophysiology, and mechanisms. *Am. Heart J.* **149**, 33–45 (2005).
51. Gavrilova, O. et al. Surgical implantation of adipose tissue reverses diabetes in lipoatrophic mice. *J. Clin. Invest.* **105**, 271–278 (2000).
52. Kim, J. K., Gavrilova, O., Chen, Y., Reitman, M. L. & Shulman, G. I. Mechanism of insulin resistance in A-ZIP/F-1 fatless mice. *J. Biol. Chem.* **275**, 8456–8460 (2000).
53. Garg, A. & Misra, A. Lipodystrophies: rare disorders causing metabolic syndrome. *Endocrinol. Metab. Clin. North Am.* **33**, 305–331 (2004).
54. Danforth, E. Failure of adipocyte differentiation causes type II diabetes mellitus? *Nature Genet.* **26**, 13 (2000).
55. Miyazaki, Y. et al. Effect of pioglitazone on abdominal fat distribution and insulin sensitivity in type 2 diabetic patients. *J. Clin. Endocrinol. Metab.* **87**, 2784–2791 (2002).
56. Wilson, P. W., D'Agostino, R. B., Parise, H., Sullivan, L. & Meigs, J. B. Metabolic syndrome as a precursor of cardiovascular disease and type 2 diabetes mellitus. *Circulation* **112**, 3066–3072 (2005).
57. Laaksonen, D. E. et al. Metabolic syndrome and development of diabetes mellitus: application and validation of recently suggested definitions of the metabolic syndrome in a prospective cohort study. *Am. J. Epidemiol.* **156**, 1070–1077 (2002).
58. Stern, M. P., Williams, K. & Haffner, S. M. Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test? *Ann. Intern. Med.* **136**, 575–581 (2002).
59. Klein, B. E., Klein, R. & Lee, K. E. Components of the metabolic syndrome and risk of cardiovascular disease and diabetes in beaver dam. *Diabetes Care* **25**, 1790–1794 (2002).
60. Macchia, A. et al. A clinically practicable diagnostic score for metabolic syndrome improves its predictivity of diabetes mellitus: the Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto miocardico (GISSI)-Prevenzione scoring. *Am. Heart J.* **151**, 754.e7–754.e17 (2006).
61. UK Prospective Diabetes Study (UKPDS) Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). *Lancet* **352**, 854–865 (1998).
62. Libby, P. & Plutzky, J. Diabetic macrovascular disease: the glucose paradox? *Circulation* **106**, 2760–2763 (2002).
63. Laakso, M. Hyperglycemia and cardiovascular disease in type 2 diabetes. *Diabetes* **48**, 937–942 (1999).
64. Grundy, S. M. et al. Definition of metabolic syndrome. Report of the National Heart, Lung and Blood Institute/American Heart Association conference on scientific issues related to definition. *Circulation* **109**, 433–438 (2004).
65. Golden, S. H. et al. Risk factor groupings related to insulin resistance and their synergistic effects on subclinical atherosclerosis: the atherosclerosis risk in communities study. *Diabetes* **51**, 3069–3076 (2002).
66. Wilson, P. W. et al. Prediction of coronary heart disease using risk factor categories. *Circulation* **97**, 1837–1847 (1998).
67. Wannamethee, S. G., Shaper, A. G., Lennon, L. & Morris, R. W. Metabolic syndrome vs Framingham Risk Score for prediction of coronary heart disease, stroke, and type 2 diabetes mellitus. *Arch. Intern. Med.* **165**, 2644–2650 (2005).
68. Stern, M. P., Williams, K., Gonzalez-Villalpando, C., Hunt, K. J. & Haffner, S. M. Does the metabolic syndrome improve identification of individuals at risk of type 2 diabetes and/or cardiovascular disease? *Diabetes Care* **27**, 2676–2681 (2004).
69. Assmann, G., Cullen, P. & Schulte, H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Munster (PROCAM) study. *Circulation* **105**, 310–315 (2002).
70. Girman, C. J. et al. The metabolic syndrome and risk of major coronary events in the Scandinavian Simvastatin Survival Study (4S) and the Air Force/Texas Coronary Atherosclerosis Prevention Study (AFCAPS/TexCAPS). *Am. J. Cardiol.* **93**, 136–141 (2004).
71. Lamarche, B. et al. Fasting insulin and apolipoprotein B levels and low-density lipoprotein particle size as risk factors for ischemic heart disease. *J. Am. Med. Assoc.* **279**, 1955–1961 (1998).
72. Hegele, R. A. Insulin resistance in human partial lipodystrophy. *Curr. Atheroscler. Rep.* **2**, 397–404 (2000).
73. Barrett-Connor, E. & Khaw, K. T. Cigarette smoking and increased central adiposity. *Ann. Intern. Med.* **111**, 783–787 (1989).
74. Shimokata, H., Muller, D. C. & Andres, R. Studies in the distribution of body fat. III. Effects of cigarette smoking. *J. Am. Med. Assoc.* **261**, 1169–1173 (1989).
75. Reaven, G. & Tsao, P. S. Insulin resistance and compensatory hyperinsulinemia: the key player between cigarette smoking and cardiovascular disease? *J. Am. Coll. Cardiol.* **41**, 1044–1047 (2003).
76. Pouliot, M. C. et al. Visceral obesity in men. Associations with glucose tolerance, plasma insulin, and lipoprotein levels. *Diabetes* **41**, 826–834 (1992).
77. Després, J. P., Lemieux, I. & Prud'homme, D. Treatment of obesity: need to focus on high risk abdominally obese patients. *Brit. Med. J.* **322**, 716–720 (2001).
78. Jorgensen, M. E. et al. Obesity and central fat pattern among Greenland Inuit and a general population of Denmark (Inter99): relationship to metabolic risk factors. *Int. J. Obes. Relat. Metab. Disord.* **27**, 1507–1515 (2003).

79. Ross, R. *et al.* Reduction in obesity and related comorbid conditions after diet-induced weight loss or exercise-induced weight loss in men. A randomized, controlled trial. *Ann. Intern. Med.* **133**, 92–103 (2000).
80. Orchard, T. J. *et al.* The effect of metformin and intensive lifestyle intervention on the metabolic syndrome: the Diabetes Prevention Program randomized trial. *Ann. Intern. Med.* **142**, 611–619 (2005).
81. Katzmarzyk, P. T. *et al.* Targeting the metabolic syndrome with exercise: evidence from the HERITAGE Family Study. *Med. Sci. Sports Exerc.* **35**, 1703–1709 (2003).
82. Dumont, M., Mauriège, P., Bergeron, J., Després, J. P. & Prud'homme, D. Effect of a six month gemfibrozil treatment and dietary recommendations on the metabolic risk profile of visceral obese men. *Int. J. Obes. Relat. Metab. Disord.* **25**, 1136–1143 (2001).
83. Leon, A. S. & Sanchez, O. A. Response of blood lipids to exercise training alone or combined with dietary intervention. *Med. Sci. Sports Exerc.* **33**, S502–S515 (2001).
84. Lemieux, I. *et al.* Waist girth does not predict metabolic complications in severely obese men. *Diabetes Care* **29**, 1417–1419 (2006).
85. Lemieux, I. *et al.* Hypertriglyceridemic waist. A marker of the atherogenic metabolic triad (hyperinsulinemia, hyperapolipoprotein B, small, dense LDL) in men? *Circulation* **102**, 179–184 (2000).
86. Underwood, P. M. Cardiovascular risk, the metabolic syndrome and the hypertriglyceridaemic waist. *Curr. Opin. Lipidol.* **15**, 495–497 (2004).
87. Kahn, H. S. & Valdez, R. Metabolic risks identified by the combination of enlarged waist and elevated triacylglycerol concentration. *Am. J. Clin. Nutr.* **78**, 928–934 (2003).
88. Lean, M. E., Han, T. S. & Morrison, C. E. Waist circumference as a measure for indicating need for weight management. *Br. Med. J.* **311**, 158–161 (1995).
89. Zhu, S. *et al.* Waist circumference and obesity-associated risk factors among whites in the third National Health and Nutrition Examination Survey: clinical action thresholds. *Am. J. Clin. Nutr.* **76**, 743–749 (2002).
90. Beckley, E. T. New ADA initiative moves beyond 'metabolic syndrome'. 'Cardiometabolic risk' proposed as umbrella term for diabetes risk factors. *DOC News* **3**, 1–3 (2006).
91. Eckel, R. H., Kahn, R., Robertson, R. M. & Rizza, R. A. Preventing cardiovascular disease and diabetes: a call to action from the American Diabetes Association and the American Heart Association. *Circulation* **113**, 2943–2946 (2006).
92. Després, J. P. From CVD risk to cardiometabolic risk. Presented at the Symposium of the International Chair on Cardiometabolic Risk. XIV International Symposium on Atherosclerosis (Rome, 18–22 June 2006).
93. Després, J. P. Visceral adiposity. Presented at the 66th Scientific Sessions, American Diabetes Association (Washington, DC, 9–13 June 2006).
94. Bouchard, C. *et al.* The response to long-term overfeeding in identical twins. *N. Engl. J. Med.* **322**, 1477–1482 (1990).

Acknowledgements The work of the authors has been supported by research grants from the Canadian Institutes of Health Research, the Canadian Diabetes Association, the Heart and Stroke Foundation and by the Foundation of the Québec Heart Institute. J.-P.D. is Scientific Director of the International Chair on Cardiometabolic Risk, which is supported by an unrestricted grant from Sanofi Aventis awarded to Université Laval.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to J.-P.D. (jean-pierre.despres@crhl.ulaval.ca).

A Mesozoic gliding mammal from northeastern China

Jin Meng^{1,2}, Yaoming Hu^{2†}, Yuanqing Wang², Xiaolin Wang² & Chuankui Li²

Gliding flight has independently evolved many times in vertebrates. Direct evidence of gliding is rare in fossil records and is unknown in mammals from the Mesozoic era. Here we report a new Mesozoic mammal from Inner Mongolia, China, that represents a previously unknown group characterized by a highly specialized insectivorous dentition and a sizable patagium (flying membrane) for gliding flight. The patagium is covered with dense hair and supported by an elongated tail and limbs; the latter also bear many features adapted for arboreal life. This discovery extends the earliest record of gliding flight for mammals to at least 70 million years earlier in geological history, and demonstrates that early mammals were diverse in their locomotor strategies and lifestyles; they had experimented with an aerial habit at about the same time as, if not earlier than, when birds endeavoured to exploit the sky.

Systematic palaeontology

Mammalia Linnaeus, 1758

Volaticotheria ord. nov.

Volaticotheriidae fam. nov.

Volaticotherium antiquus gen. et sp. nov.

Etymology. *Volaticus* (Latin): winged, flying; *theri* (Greek): beast; *antiquus* (Latin): ancient.

Holotype. A squashed skeleton preserved in the part and counterpart of a split slab that also contains numerous carapace valves of small conchostracans (*Euestheria*)¹ (Fig. 1). Some parts of the counterpart are missing. The specimen is housed in the Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing, China (IVPP V14739; see Supplementary Information).

Horizon and locality. Daohugou, Ningcheng County, Inner Mongolia, China. Age estimates of the Daohugou beds are controversial, ranging from Middle Jurassic², through Late Jurassic³, to Late Jurassic/Early Cretaceous^{4,5}. However, the Daohugou fauna is older than the Jehol Biota that dates back to 125 Myr ago⁶ (see Supplementary Information).

Diagnosis. A small squirrel-sized mammal that differs from all known Mesozoic mammals in having the following features: presence of a sizable furry patagium; elongated limb bones; femur with a small, ovoid-shaped head that lacks a neck; hallux diverging medially; proximal phalanges with pronounced flexor-sheath ridges; long tail with elongated and dorsoventrally flat caudal vertebrae and haemal arches; a highly differentiated dentition with a dental formula of I³-C¹-P⁴-M³/I²-C¹-P⁴-M²; incisors small and conical; canines long and sharp; molariforms with tall, sharp and posteriorly recumbent cusps that are aligned in line but are deeply separated; cusps of lower molariforms are more posteriorly recumbent and tightly packed than the uppers; cusp d of M₁ labially overlapping cusp b of M₂ (see Supplementary Information for a detailed description).

Comparison and affinity

In addition to its unique patagium and related skeletal features, the cheek teeth of *Volaticotherium antiquus* (Fig. 1) differ from those of any known Mesozoic mammals except that they resemble, and are

probably derivable from, those of triconodonts because the tooth cusps align anteroposteriorly. The teeth are, however, more specialized and differentiated than those of any known triconodont. In addition to the high, posteriorly recumbent cusps on molariforms, the cheek teeth of *V. antiquus* differ from triconodont teeth in having cusp d of M₁ labially overlapping cusp b of M₂, and in lacking both an anterior accessory cuspace on the lower premolariforms and the interlock relationship of molariforms. Moreover, the presence of a posteriorly positioned angular process on the dentary further distinguishes *V. antiquus* from triconodonts.

The upper molariforms of *V. antiquus* are most comparable with two isolated teeth, identified as lower molars, of *Ichthyoconodon* from the Early Cretaceous littoral sediments, Morocco⁷. *Ichthyoconodon* is currently treated as a eutriconodontan, but its affinities are problematic^{8,9}. If the known teeth of *Ichthyoconodon* are actually upper ones, they differ from those of *V. antiquus* by having lower cusps that are less recumbent posteriorly and more trenchant than those of *V. antiquus*. The lower molariforms of *V. antiquus* differ from those of any known triconodonts.

V. antiquus has body hair and a single-boned mandible, which are conventionally regarded as typical mammalian features. A single-boned mandible and the presence of paired sesamoid bones between the metatarsus and phalange further differentiate *Volaticotherium* from *Sinoconodon*, *Morganucodon* and docodontans^{9,10}. The presence of an internasal process of the premaxilla is a primitive feature in Mammalia, but in *V. antiquus* the process does not extend posteriorly to contact the nasals, thus differing from the condition in *Sinoconodon*, *Morganucodon* and non-mammalian cynodonts. In addition, the naris is proportionally larger than those of *Sinoconodon*, *Morganucodon* and non-mammalian cynodonts.

A phylogenetic analysis of selected Mesozoic mammals based on a data set consisting of 58 taxa and 435 characters identifies *Volaticotherium* as an independent clade in Mammalia, defined either as the crown group¹¹ or as a more inclusive group⁹ (Fig. 2; Supplementary Information). Within Mammalia, *Volaticotherium* forms the sister taxon to the clade containing eutriconodontans, multituberculates and trechnotherians. The phylogenetic position and numerous unique features indicate that *Volaticotherium* represents a previously unknown and highly specialized mammalian clade that

¹Division of Paleontology, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA. ²Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, PO Box 643, Beijing 100044, China. †Present address: Department of Anatomical Sciences, Stony Brook University, Stony Brook, New York 11794-8081, USA.

was probably derived from a stock with triconodont-like dentitions. We recognize this clade as a new mammalian Order.

Patagium and supporting structures

The most prominent feature of *Volaticotherium antiquus* is the patagium that is covered with dense hair. The soft structures are preserved

as impressions (Figs 1a, 3a; Supplementary Information). The patagium impression is not completely exposed but is sufficient to indicate its large size, certainly larger than that of the Oligocene gliding rodent¹². In some areas, perpendicular orientations of the hair on different layers of impressions suggest folding of the patagium in preservation (Fig. 3a). The patagium impression continues with skin

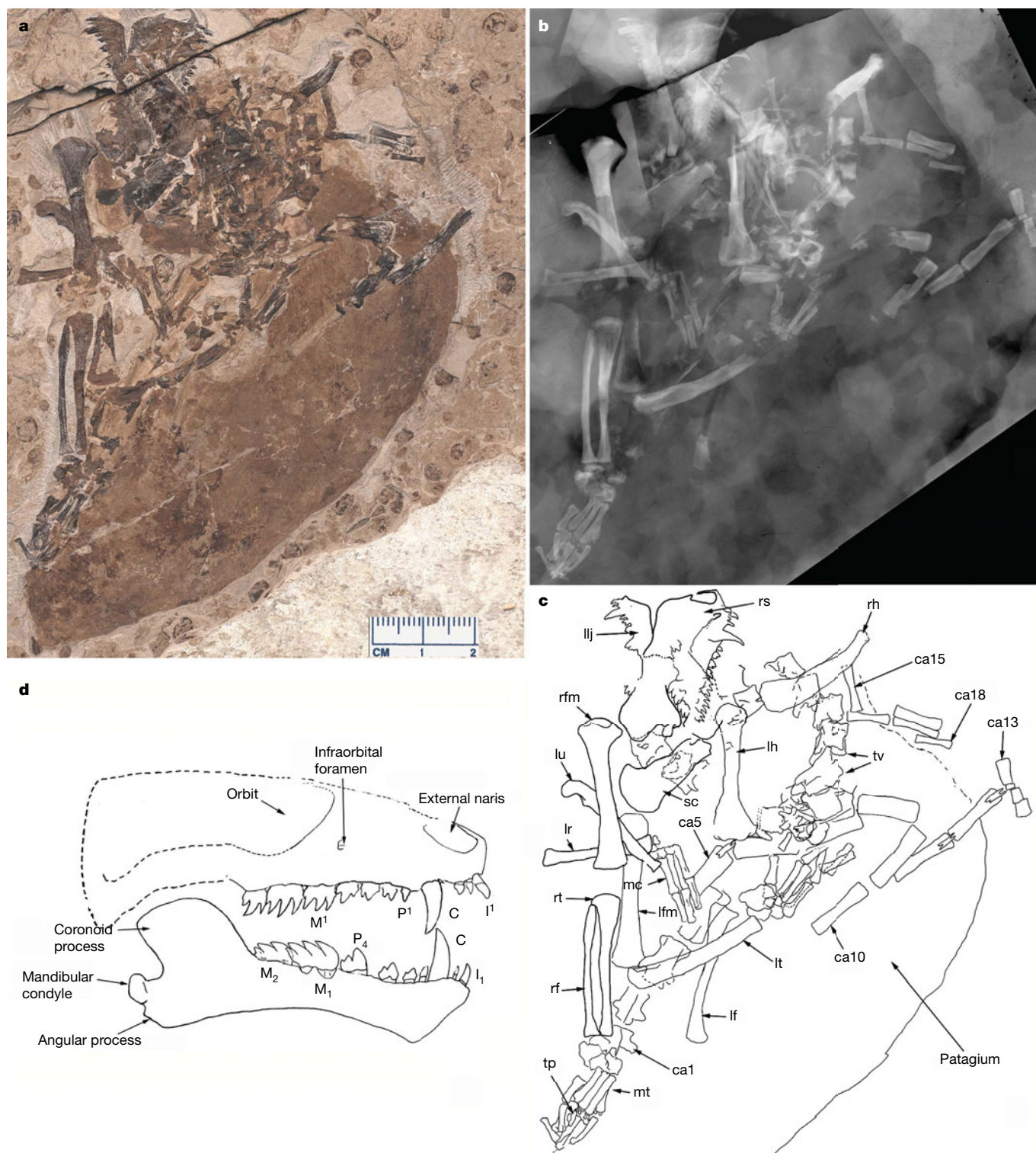


Figure 1 | *Volaticotherium antiquus* gen. et sp. nov. **a**, **b**, The specimen of *V. antiquus* (**a**) and its X-radiographic image (**b**; holotype, IVPP V14739). **c**, Line drawing illustrating major skeletal elements in **a** and **b**. ca1, 5, 10, 13, 15, 18 are the 1st, 5th, 10th, 13th, 15th and 18th preserved caudal vertebrae, respectively; lf, left fibula; lfm, left femur; lh, left humerus; llj, left lower jaw;

lr, left radius; lt, left tibia; lu, left ulna; mc, metacarpals; mt, metatarsals; rf, right fibula; rfm, right femur; rh, right humerus; rs, rostrum of the skull; rt, right tibia; sc, scapula; tp, terminal phalanx; tv, thoracic vertebrae. **a–c** are the same scale. **d**, Reconstructed skull and lower jaws of *V. antiquus*.

and hair impressions surrounding the area where the skeletal elements are concentrated; it stretches to the phalange area of the foot and sandwiches the tarsal region. These suggest that the limbs, probably including the pes and manus, had participated in supporting the patagium. The hair on the patagium and surrounding the body is fine and even, but that along several proximal caudal vertebrae is relatively thick. The parallel orientations of hair along some caudal vertebrae are indicative of the original anatomical relationships of the hair, skin and bones.

The furry patagium of *Volaticotherium antiquus* is direct evidence of its adaptation to gliding flight. Animal flight requires a high degree of morphological adaptation, of which the most important feature is the development of a sufficient surface area to form an airfoil that supports the weight of the animal and generates lift for the animal to travel horizontally in air. For mammals, such a surface area is obtained by the development of the patagium, a fold of skin membrane from the body, which usually stretches between the fore and hind legs^{13–15} and is controlled by complex muscles¹⁶. To support an

increased area of patagium, limbs of extant gliding mammals are commonly elongated^{17,18}.

Mesozoic mammals generally have relatively short limbs, whether they were terrestrial, such as *Jeholodens*¹⁹ and *Zhangheotherium*²⁰, or tree climbers, such as *Eomaia*²¹ and *Sinodelphys*²². In contrast, all limb elements of *V. antiquus* are significantly longer than the corresponding elements in other Mesozoic mammals (see Supplementary Information). Given its phylogenetic relationship, *V. antiquus* was probably derived from a short-limbed ancestor. The ratio of limb lengths of *V. antiquus* to those of other Mesozoic mammals is similar to that between living gliding and non-gliding mammals. In addition to limb elongation, the caudal vertebrae of *V. antiquus* are long, broad and dorsoventrally flat with a low, dorsal, longitudinal ridge (Fig. 3c). The haemal arch that ventrally bridges two adjacent caudal vertebrae is also long and flat. Several caudal vertebrae are also sandwiched by the impressions of the patagium. These structures and their relationships with the patagium suggest the participation of the proximal caudal vertebrae in supporting the patagium and thus

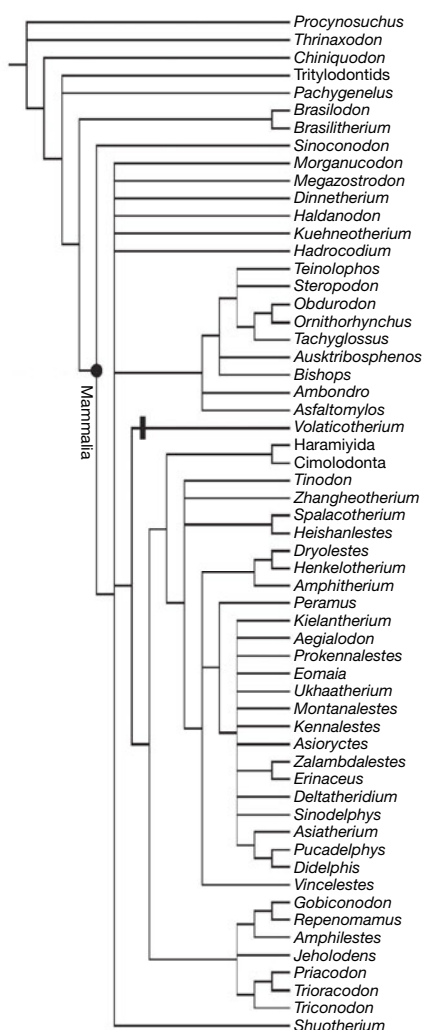


Figure 2 | Phylogenetic relationships of major Mesozoic mammalian groups, including *Volaticotherium antiquus*. The cladogram (tree length = 1,520; consistency index (CI) = 0.408; retention index (RI) = 0.609) is the strict consensus of 72 equally most parsimonious trees (tree length = 1,354; CI = 0.458; RI = 0.749) obtained from 1,000 replications of random heuristic searches using PAUP 4.0b10 (Mac version)⁴³ with a parsimony principle assumed and all characters unordered. The data matrix consists of 58 taxa and 435 characters (see Supplementary Information); it was constructed using Mesquite (v. 1.06)⁴⁴ (<http://mesquiteproject.org>) and converted into the final data matrix using MacClade 4.08 for Mac OS X⁴⁵.



Figure 3 | Detailed features related to gliding and arboreal locomotion of *Volaticotherium antiquus*. **a**, Close-up view of the patagium impression of *V. antiquus* (IVPP V14739), with white arrows indicating nearly perpendicular orientations of hair on two layers of impressions. **b**, Ventral views of two long, flat, caudal vertebrae with white arrows pointing to two long, flat, haemal arches. **c**, Dorsal (anterior) and proximal views of the right femur, showing the small, ovoid-shaped head that lacks a neck. **d**, Comparison of the femoral head of *V. antiquus* (labelled) with that of a morganucodontid (modified from ref. 25). fh, femoral head, gt, greater trochanter; lt, lesser trochanter.

presence of uropatagia between the tail and hind limbs. The morphologies of caudal vertebrae and haemal arches also indicate a stiff tail that had limited lateral and dorsoventral movements, and could act as a stabilizer when the animal was gliding, consistent with the hypothesis that longitudinal control and stability of a glider are achieved more easily with a long, dorsoventrally flattened tail^{14,15,23}.

Other gliding and arboreal features

Two hind limbs are preserved with all elements nearly in original articulations (Fig. 1; Supplementary Information). The femur has a small, oval-shaped head and lacks a femoral neck (Fig. 3c, d), which is unique among mammals⁷. Although the hip structure was not preserved, the shape of the femoral head alone reveals that the acetabulum is probably shallow and open, perhaps similar to those of arboreal didelphids²⁴. Unlike the ball-and-socket hip joint that provides ample excursion for legs in other mammals²⁵, the peculiar femoral head of *Volaticotherium* must have restricted flexibility of the leg in rotational movements, but could have allowed the leg to be extended laterally and remain steady during a glide.

Many other limb features of *V. antiquus* are related to arboreal locomotion (Fig. 1; Supplementary Information), which is essential for *V. antiquus* to obtain altitude to a launch point for a glide. The hallucal metatarsal diverges medially at a 35° angle from the second metatarsal. A sesamoid bone is at the distal phalange joint of the hallux. Pedal metatarsals and proximal phalanges are dorsally arched with paired sesamoid bones between them. The proximal phalange is relatively long compared to the metatarsal and has pronounced flexor sheath ridges that flare outward to confine a longitudinal groove on the ventral surface of the element. The proximal end of the intermediate phalange is dorsoventrally deep, corresponding to a robust and well-trochleated distal end of the proximal phalange. The terminal hallucal phalange, the only preserved terminal phalange, is large and mediolaterally compressed, indicating a large, deep claw. Only four digits of the manus are partially preserved; they are structurally similar to those of the pes. These features are significantly different from those that have a more generalized locomotion adaptation, such as morganucodontids²⁵, eutriconodontans^{19,26} and symmetrodontans²⁰, but are similar to various arboreal and gliding mammals, such as didelphids, *Carpolestes*, paromomyids, *Cynocephalus* and gliding squirrels, as a result of adaptation for vertical postures, climbing and other arboreal behaviours^{24,27–31}. The pedal and manual morphologies of *V. antiquus* show that both the hind and fore feet are relatively long, and could have a strong ability to flex the digits to secure grips around branches that are large relative to the size of the animal. When not committing its distal limbs to the patagial lifting surface, the hands and feet of *V. antiquus* could be used for surface transport and stable quadrupedal landing, as in some living, gliding mammals³². Coupled with the unique joint between the femur and hip, these morphologies indicate an extreme arboreal animal, a tree dweller that could forage on trees, similarly to some extant gliding mammals³³.

Life style

Extant gliding mammals have a body mass range of 10 g–1.5 kg, broadly similar to that of bats^{13,18,23,34}. Given its 28.3 mm jaw length, the skull length of *V. antiquus* is estimated to be about 35 mm and the combined head and body length around 120–140 mm with a body mass of around 70 g, similar to that of the flying squirrel *Glaucomys volans*¹⁸. *Volaticotherium antiquus* was most probably nocturnal, not only because small Mesozoic mammals are generally thought to be nocturnal^{9,35}, but also because gliding mammals are predominantly arboreal and nocturnal^{13,18,23,36}. *Volaticotherium antiquus* could have been an agile glider because of its low patagial load, resulting from its small body mass and relatively large patagium, but it probably did not have the ability to habitually capture airborne prey, as many insectivorous bats do. The patagium of *V. antiquus* has a low aspect ratio, as indicated by its limb lengths, which would have restricted it to gliding

flight for transport and meant it was not manoeuvrable enough for aerial foraging.

Gliding flight is thought to be ancestral to bats^{33,37–40} and arose up to seven times independently within three other mammalian groups: Marsupialia, Dermoptera and Rodentia^{13,14,17,23,36,41}. Gliding as a locomotor strategy has its advantages, such as being energetically economic, allowing a wider foraging area and enabling evasion from predators in arboreal environments^{15,23,33,42}, but fossils of gliding mammals are extremely rare. The earliest confirmed record of bats dates from the Early Eocene about 51 Myr ago⁴⁰; whereas the earliest confirmed gliding mammal with a small patagium is a Late Oligocene eomyid rodent¹². There was no convincing evidence of flight adaptation in mammals before 51 Myr ago. The morphological evidence for the earliest gliding locomotion from *V. antiquus* shows that mammals had experimented with aerial life much earlier than previously expected: probably at the same time as, if not earlier than, when birds exploited the sky. The absence of identifiable gliding mammals in the fossil record during a period of at least 70 million years is probably due to poor preservation of these small animals, particularly their characteristic gliding structures, such as the patagium.

Received 22 May; accepted 4 September 2006.

- Shen, Y.-b., Chen, P.-j. & Huang, D.-y. Age of fossil conchostracans from Daohugou of Ningcheng, Inner Mongolia [in Chinese with English abstract]. *J. Stratigr.* **27**, 311–313 (2003).
- Chen, W. *et al.* Isotope geochronology of the fossil-bearing beds in the Daohugou area, Ningcheng, Inner Mongolia [in Chinese]. *Geol. Bull. Chin.* **23**, 1165–1169 (2004).
- Zhang, J.-f. Discovery of Daohugou Biota (Pre-Jehol Biota) with a discussion on its geological age [in Chinese with English abstract]. *J. Stratigr.* **26**, 173–177 (2002).
- He, H.-y. *et al.* ⁴⁰Ar/³⁹Ar dating of ignimbrite from Inner Mongolia, northeastern China, indicates a post-Middle Jurassic age for the overlying Daohugou Bed. *Geophys. Res. Lett.* **31**, L20609, doi:10.1029/2004GL020792 (2004).
- Wang, X.-l. *et al.* Stratigraphy and age of the Daohugou bed in Ningcheng, Inner Mongolia. *Chin. Sci. Bull.* **50**, 2369–2376 (2005).
- Swisher, C. C. III *et al.* Cretaceous age for the feathered dinosaurs of Liaoning, China. *Nature* **400**, 58–61 (1999).
- Sigogneau-Russell, D. Two possibly aquatic triconodont mammals from the Early Cretaceous of Morocco. *Acta Palaeontol. Pol.* **40**, 149–162 (1995).
- Cifelli, R. L., Wible, J. R. & Jenkins, F. A. Jr. Triconodont mammals from the Cloverly Formation (Lower Cretaceous), Montana and Wyoming. *J. Vert. Paleontol.* **18**, 237–241 (1998).
- Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z.-x. *Mammals from the Age of Dinosaurs: Origins, Evolution and Structure* (Columbia Univ. Press, New York, 2004).
- Ji, Q., Luo, Z.-x., Yuan, C.-x. & Tabrum, A. R. A swimming mammaliaform from the Middle Jurassic and ecomorphological diversification of early mammals. *Science* **311**, 1123–1127 (2006).
- Rowe, T. B. Definition, diagnosis, and origin of Mammalia. *J. Vert. Paleontol.* **8**, 241–264 (1988).
- Storch, G., Engesser, B. & Wuttke, M. Oldest fossil record of gliding in rodents. *Nature* **379**, 439–441 (1996).
- Rayner, J. M. V. Flight adaptations in vertebrates. *Symp. Zool. Soc. Lond.* **48**, 137–172 (1981).
- Norberg, U. M. in *Functional Vertebrate Morphology* (eds Hildebrand, M., Bramble, D. M., Liem, K. F. & Wake, D. B.) 129–158 (The Belknap Press, Harvard Univ., Massachusetts, 1985).
- Norberg, U. M. Evolution of vertebrate flight: an aerodynamic model for the transition from gliding to active flight. *Am. Nat.* **126**, 303–327 (1985).
- Johnson-Murray, J. Myology of the gliding membranes of some Petauristine rodents (genera: *Glaucomys*, *Pteromys*, *Petinomys*, and *Petaurista*). *J. Mamm.* **58**, 374–384 (1977).
- Runestad, J. A. & Ruff, C. B. Structural adaptations for gliding mammals with implications for locomotor behaviour in paromomyids. *Am. J. Phys. Anthropol.* **98**, 101–119 (1995).
- Thorigton, R. W. Jr & Heaney, L. R. Body proportions and gliding adaptations of flying squirrels. *J. Mamm.* **62**, 101–114 (1981).
- Ji, Q., Luo, Z. & Ji, S. A Chinese triconodont mammal and mosaic evolution of the mammalian skeleton. *Nature* **398**, 326–330 (1999).
- Hu, Y., Wang, Y., Luo, Z. & Li, C. A new symmetrodont mammal from China and its implications for mammalian evolution. *Nature* **390**, 137–142 (1997).
- Ji, Q. *et al.* The earliest known eutherian mammal. *Nature* **416**, 816–822 (2002).
- Luo, Z.-x., Ji, Q., Wible, J. R. & Yuan, C.-x. An Early Cretaceous tribosphenic mammal and metatherian evolution. *Science* **302**, 1934–1940 (2003).
- Lindenmayer, D. *Gliders of Australia: A Natural History* (Univ. of New South Wales Press, New South Wales, 2002).

24. Argot, C. Functional-adaptive analysis of the hindlimb anatomy of extant marsupials and the paleobiology of the Paleocene marsupials *Mayulestes ferox* and *Pucadelphys andinus*. *J. Morphol.* **253**, 76–108 (2002).
25. Jenkins, F. A. Jr & Parrington, F. R. The postcranial skeletons of the Triassic mammals *Eozostrodon*, *Megazostrodon* and *Erythrotherium*. *Phil. Trans. R. Soc. Lond. B* **273**, 387–431 (1976).
26. Hu, Y., Meng, J., Wang, Y. & Li, C. Large Mesozoic mammals fed on young dinosaurs. *Nature* **433**, 149–153 (2005).
27. Beard, K. C. Gliding behaviour and palaeoecology of the alleged primate family Paromomyidae (Mammalia, Dermoptera). *Nature* **345**, 340–341 (1990).
28. Beard, K. C. (1993). in *Primates and Their Relatives in Phylogenetic Perspective* (ed. MacPhee, R. D. E.) 63–90 (Plenum, New York, 1993).
29. Hamrick, M. W., Rosenman, B. A. & Brush, J. A. Phalangeal morphology of the Paromomyidae (?Primates, Plesiadapiformes): the evidence for gliding behavior reconsidered. *Am. J. Phys. Anthropol.* **109**, 397–413 (1999).
30. Szalay, F. S. & Lucas, S. G. in *Primates and Their Relatives in Phylogenetic Perspective* (ed. MacPhee, R. D. E.) 187–226 (Plenum, New York, 1993).
31. Bloch, J. I. & Boyer, D. M. Grasping primate origins. *Science* **298**, 1606–1610 (2002).
32. Caple, G., Balda, R. P. & Willis, W. R. The physics of leaping animals and the evolution of preflight. *Am. Nat.* **121**, 455–476 (1983).
33. Stafford, B. J., Thorington, R. W. Jr & Kawamichi, T. Positional behavior of Japanese giant flying squirrels (*Petaurista leucogenys*). *J. Mamm.* **84**, 63–271 (2003).
34. Norberg, U. M. Vertebrate flight: mechanics, physiology, morphology, ecology and evolution. *Zoophysiol.* **27**, 1–291 (1990).
35. Lillegraven, J. A., Kielan-Jaworowska, Z. & Clemens, W. A. (eds) *Mesozoic Mammals: The First Two-Thirds of Mammalian History* (Univ. of California Press: Berkeley, 1979).
36. Nowak, R. M. *Walker's Mammals of the World* 6th edn (Johns Hopkins Univ. Press, Baltimore, 1999).
37. Norberg, U. M. in *European Bat Research* (eds Hanák, V., Horáček, I. & Gaisler, J.) 197–211 (Charles Univ. Press, Praha, 1989).
38. Fenton, M. B., Audet, D., Obrist, M. K. & Rydell, J. Signal strength, timing, and self-deafening: the evolution of echolocation in bats. *Paleobiology* **21**, 229–242 (1995).
39. Simmons, N. B. & Geisler, J. H. Phylogenetic relationships of *Icaronycteris*, *Archaeonycteris*, *Hassianycteris*, and *Palaeochiropteryx* to extant bat lineages, with comments on the evolution of echolocation and foraging strategies in Microchiroptera. *Bull. Am. Mus. Nat. Hist.* **235**, 1–182 (1998).
40. Gunnell, G. F. & Simmons, N. B. Fossil evidence and the origin of bats. *J. Mamm. Evol.* **12**, 209–246 (2005).
41. Thorington, R. W. Jr. Flying squirrels are monophyletic. *Science* **225**, 1048–1050 (1984).
42. Jackson, S. M. Glide angle in the genus *Petaurus* and a review of gliding in mammals. *Mamm. Rev.* **30**, 9–30 (1999).
43. Swofford, D. L. *PAUP*. Phylogenetic Analysis Using Parsimony and Other Methods. Version 4* (Sinauer Assoc., Sunderland, Massachusetts, 2002).
44. Maddison, W. P. & Maddison, D. R. *Mesquite: a Modular System for Evolutionary Analysis. Version 1.06* (<http://mesquiteproject.org>) 2005a).
45. Maddison, W. P. & Maddison, D. R. *MacClade: Analysis of Phylogeny and Character Evolution. Version 4.08* (Sinauer Assoc., Sunderland, Massachusetts, 2005b).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Z. Zhou, F. Zhang, X. Xu, Y. Wang, F. Jin, and J. Zhang for help in coordinating the research and fieldwork; S. Xie for specimen preparation; M. Yang for drawing Supplementary Fig. 13; and L. Meeker for X-radiography. The research has been funded by the Chinese Academy of Sciences, the National Natural Science Foundation of China; and Major Basic Research Projects of the Ministry of Science and Technology, China. The research of Y.H. is also supported by a fellowship from the American Museum of Natural History, New York.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.M. (jmeng@amnh.org) or Y.H. (huyaoming@ivpp.ac.cn).

ARTICLES

An *SCN9A* channelopathy causes congenital inability to experience pain

James J. Cox^{1*}, Frank Reimann^{2*}, Adeline K. Nicholas¹, Gemma Thornton¹, Emma Roberts³, Kelly Springell³, Gulshan Karbani⁴, Hussain Jafri⁵, Jovaria Mannan⁶, Yasmin Raashid⁷, Lihadh Al-Gazali⁸, Henan Hamamy⁹, Enza Maria Valente¹⁰, Shaun Gorman¹¹, Richard Williams¹², Duncan P. McHale¹², John N. Wood¹³, Fiona M. Gribble² & C. Geoffrey Woods¹

The complete inability to sense pain in an otherwise healthy individual is a very rare phenotype. In three consanguineous families from northern Pakistan, we mapped the condition as an autosomal-recessive trait to chromosome 2q24.3. This region contains the gene *SCN9A*, encoding the α -subunit of the voltage-gated sodium channel, Na_v1.7, which is strongly expressed in nociceptive neurons. Sequence analysis of *SCN9A* in affected individuals revealed three distinct homozygous nonsense mutations (S459X, I767X and W897X). We show that these mutations cause loss of function of Na_v1.7 by co-expression of wild-type or mutant human Na_v1.7 with sodium channel β_1 and β_2 subunits in HEK293 cells. In cells expressing mutant Na_v1.7, the currents were no greater than background. Our data suggest that *SCN9A* is an essential and non-redundant requirement for nociception in humans. These findings should stimulate the search for novel analgesics that selectively target this sodium channel subunit.

Pain is an essential sense that has evolved in all complex organisms to minimize tissue and cellular damage, and hence prolong survival. The onset of pain results in the adoption of behaviours that both remove the organism from a 'dangerous environment' and allow for tissue repair; for example, resting a broken limb so that new bone can form. Pain also protects us from our environment, by teaching us what situations and behaviours are likely to lead to injury. Pain pathways operate at numerous levels in the nervous system and are under both voluntary and involuntary control. Blockade of this system with analgesics has been a major pharmacological achievement.

Whereas individuals with a congenital absence of the sense of vision or of hearing are relatively common, a congenital absence of the sense of pain is very rare. The first case of a patient with a congenital inability to perceive pain was said to have been reported in the early twentieth century¹. Only a handful of such patients have since been described and are usually categorized as having 'congenital indifference to pain' (OMIM 243000, also known as autosomal recessive congenital analgesia) or as being misdiagnoses of 'congenital insensitivity to pain' (OMIM 608654, also known as hereditary sensory and autonomic neuropathy type 5 (HSAN5))^{2–4}. Historically these two conditions have been largely distinguished by the absence or presence, respectively, of an associated neuropathy². The existence of congenital indifference to pain has, however, been questioned and the nomenclature surrounding its differentiation from congenital insensitivity to pain has been the focus of controversy^{2,3,5} (see Methods). Here we describe individuals from three families with the extraordinary phenotype of a congenital inability to perceive any form of pain, in whom all other sensory modalities were preserved and the peripheral and central nervous systems were appar-

ently otherwise intact. As the clinical description of the study individuals does not exactly match pre-existing reports of either indifference to pain or insensitivity to pain, we refer to this new syndrome as 'channelopathy-associated insensitivity to pain' and show that it is caused by loss of function of the voltage-gated sodium channel gene *SCN9A*.

Absence of pain phenotype

The index case for the present study was a ten-year-old child, well known to the medical service after regularly performing 'street theatre'. He placed knives through his arms and walked on burning coals, but experienced no pain. He died before being seen on his fourteenth birthday, after jumping off a house roof. Subsequently, we studied three further consanguineous families in which there were individuals with similar histories of a lack of pain appreciation, each originating from northern Pakistan and part of the Qureshi birdari/clan (Fig. 1). All six affected individuals had never felt any pain, at any time, in any part of their body. Even as babies they had shown no evidence of pain appreciation. None knew what pain felt like, although the older individuals realized what actions should elicit pain (including acting as if in pain after football tackles). All had injuries to their lips (some requiring later plastic surgery) and/or tongue (with loss of the distal third in two cases), caused by biting themselves in the first 4 yr of life. All had frequent bruises and cuts, and most had suffered fractures or osteomyelitis, which were only diagnosed in retrospect because of painless limping or lack of use of a limb. The children were considered of normal intelligence by their parents and teachers, and by the caring physicians. One author saw and reviewed all six affected individuals and their families. The ages at which the

¹Department of Medical Genetics, and ²Department of Clinical Biochemistry, Cambridge Institute for Medical Research, Wellcome/MRC Building, Addenbrooke's Hospital, Cambridge CB2 0XY, UK. ³Section of Ophthalmology and Neuroscience Leeds Institute of Molecular Medicine, ⁴Department of Clinical Genetics, St James's University Hospital, Leeds LS9 7TF, UK. ⁵Gene Tech Lab 146/1, Shadman Jail Road, Lahore, Pakistan. ⁶Department of Paediatrics, Fatima Jinnah Medical College, Lahore, Pakistan. ⁷Department of Obstetrics and Gynaecology, King Edward Medical University, Lahore, Pakistan. ⁸Department of Paediatrics, Faculty of Medicine and Health Sciences, United Arab Emirates University, Al-Anin P O Box 15551, United Arab Emirates. ⁹National Center for Diabetes, Endocrinology and Genetics, Amman P O Box 3165/11942, Jordan. ¹⁰IRCCS CSS, San Giovanni Rotondo and CSS Mendel, Rome I-00168, Italy. ¹¹Department of Paediatrics, St Luke's Hospital, Bradford BD5 0NA, UK. ¹²Pfizer Global Research and Development, Sandwich Laboratories, Sandwich CT13 9NJ, UK. ¹³Molecular Nociception Group, Department of Biology, University College London, London WC1E 6BT, UK.

*These authors contributed equally to this work.

children were examined and their families initially seen were (with reference to Fig. 1 and from left to right): family 1, the children were 6, 4 and 14 years old; family 2, the child was 6 years old; and family 3, the children were 12 and 10 years old.

Detailed neurological examinations revealed that each could correctly perceive the sensations of touch, warm and cold temperature, proprioception, tickle and pressure, but not painful stimuli. Pain sensation was assessed by squeezing of the Achilles' tendon, firm pressure to dorsal fingertips inflicted with a thumbnail and by venesection; all were felt but not described as painful or unpleasant. Sensation of touch was assessed by needle point and cotton wool and was normal; proprioception was normal despite the children being 'ungainly' in gross motor movements; temperature sensation was not rigorously assessed but all children could tell cold from hot in food and drink (two had received painless scalds as young children). There was no evidence of a motor or sensory neuropathy. Strength, tone, reflexes including plantar responses, and appearance of joints were all normal. Peripheral nerves were not palpably enlarged. Corneal reflex was present, and the gag reflex was reported as normal. None had symptoms of autonomic nervous system dysfunction: all sweated and flushed/blushed appropriately; there were no increased episodes of hyperpyrexia (although these did occur appropriately with infection); tongues of all children were normal with no loss of fungiform papillae; bladder control occurred within the normal age spectrum and there was no history of urinary infections, incontinence or retention; there were no episodes of unexplained vomiting or dysphagia; tear production was reported as normal, and no child had dry eyes. However, a histamine flare test and assessment of itching

was not performed. All had normal health, vision, hearing and appearance.

Nerve conduction studies of the radial nerve were conducted in the older child in family 3 and the affected individual in family 2. The results were normal with no evidence of a neuropathy, either axonal or demyelinating. Nerve biopsy of the sural nerve was performed in the same two individuals and showed a normal sensory nerve with all nerve fibre types present, of normal morphology, and in normal distribution by light and electron microscopy. One member of family 1 had a magnetic resonance imaging (MRI) brain scan, which was reported and subsequently reviewed as showing no abnormalities. Parents and other siblings reported normal pain appreciation.

Mapping the disease gene

We used a positional cloning strategy to identify the mutated gene in these families (Fig. 1). A genome-wide scan using 400 polymorphic microsatellite markers led to the identification of an 11.7-megabase (Mb) homozygous region on chromosome 2q24 shared between the affected individuals from all three families. In order to reduce the region, we sought a common haplotype between families by genotyping an additional 73 polymorphic microsatellite markers (Supplementary Table 1). This analysis failed to identify a significant shared haplotype block, suggesting that each family had a different mutation. Bioinformatics analysis of the ~50 genes in the linkage region identified *SCN9A* as the best candidate gene, and subsequent sequence analysis of *SCN9A* revealed distinct homozygous nonsense mutations in each of the three families (Fig. 2 and Supplementary Table 2). In family 1 we found a single homozygous base substitution in coding exon 15 (2691G→A, resulting in the amino acid change W897X). In family 2 we found a single homozygous base deletion of

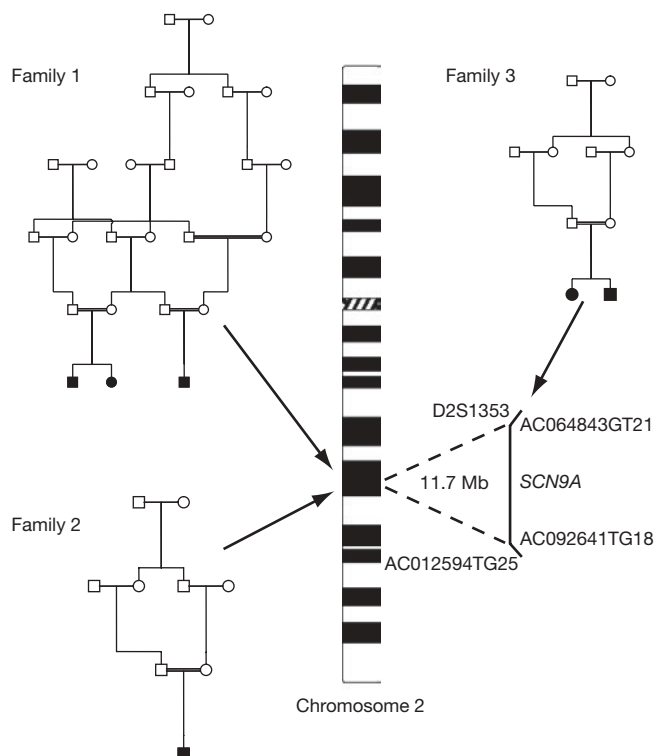


Figure 1 | The families used to map the locus for channelopathy-associated insensitivity to pain. Autozygosity mapping in families 1 and 2 on the left of the diagram enabled the identification of a shared 20 cM homozygous region on chromosome 2q24 (there were no other significant homozygous regions detected). The two-point LOD score was 3.2 at $\theta = 0$, for AC064843GT21 and AC092641TG18, but greater for more informative markers, see Supplementary Table 1. Family 3, on the right of the diagram, was used to refine this region to an 11.7-Mb shared homozygous region flanked by the heterozygous markers D2S1353 and AC012594TG25 and defined by the homozygous markers AC064843GT21 and AC092641TG18. Affected individuals are indicated with filled symbols.

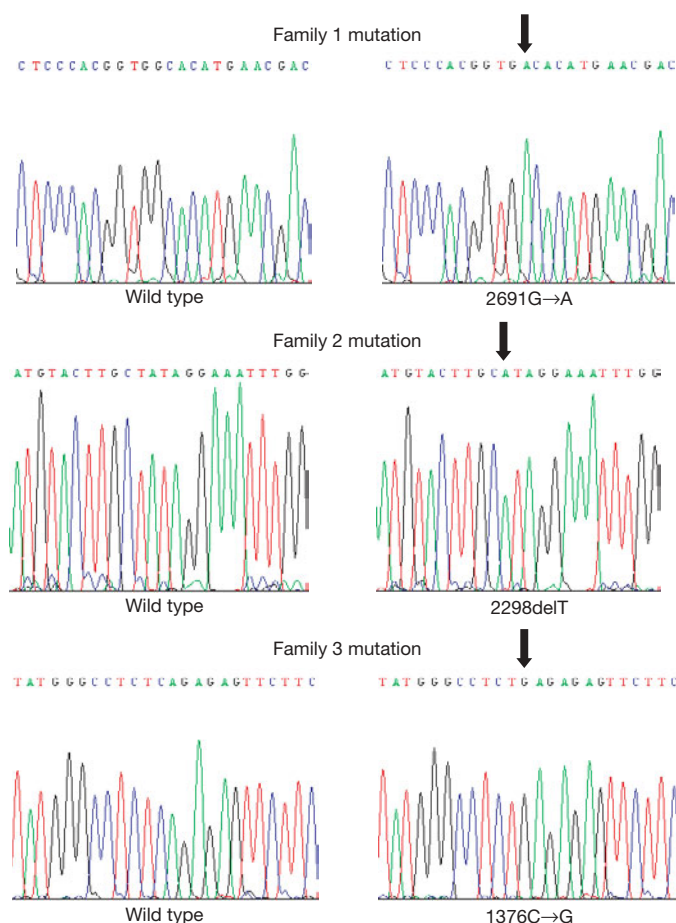


Figure 2 | Sequence chromatograms showing the mutations identified in families 1, 2 and 3. The arrows indicate the site of the mutations.

base 2298 in coding exon 13 that led to a frameshift and the amino acid change I767X. In family 3 we found a single homozygous base substitution in coding exon 10 (1376C→G, resulting in the amino acid change S459X) (Fig. 3). Each mutation was absent from 300 northern Pakistani control chromosomes and showed the expected disease segregation within the families. Although DNA was not available from the original index case, it is likely that he too had channelopathy-associated insensitivity to pain as his mother was heterozygous for the *SCN9A* nonsense mutation W897X (his father was not available to test). We considered the possibility that channelopathy-associated insensitivity to pain was inherited as a double recessive disorder in these families. However, we concluded that this was extremely unlikely because the 11.7-Mb region containing *SCN9A* was the only significant shared homozygous region between the three families and because the in-depth haplotype study of this linkage region (Supplementary Table 1) failed to identify a significant shared haplotype block.

SCN9A encodes $\text{Na}_v1.7$, the α -subunit of a tetrodotoxin-sensitive voltage-gated sodium channel that is expressed at high levels in peripheral sensory neurons, most notably in nociceptive small-diameter dorsal root ganglia (DRG) neurons^{6–8}. Voltage-gated sodium channels underlie the depolarizing phase of action potentials in excitable cells and tissue-specific expression of the various family members help to shape the excitability and repetitive firing properties of different neurons^{9,10}. The precise function of $\text{Na}_v1.7$ in sensory neurons is unclear, although immunostaining of cultured dorsal root ganglia neurons has suggested that it is targeted to the nerve terminals, where it has been proposed to have an involvement in action potential initiation^{7,11}.

Loss of function of $\text{Na}_v1.7$

The $\text{Na}_v1.7$ nonsense mutations identified in our families are expected to cause prematurely truncated proteins or nonsense-mediated messenger RNA decay and hence loss of function of $\text{Na}_v1.7$ in nociceptive neurons (Fig. 3)¹². To determine the activity of any possible stable truncated $\text{Na}_v1.7$ proteins, we carried out patch-clamp experiments in human embryonic kidney (HEK293) cells. Wild-type $\text{Na}_v1.7$, or $\text{Na}_v1.7$ containing the patient mutations, was co-expressed in HEK293 cells with the auxiliary sodium channel β_1 and β_2 subunits (encoded by *SCN1B* and *SCN2B*, respectively), which are also expressed in DRG neurons¹³ and are necessary for the normal function of voltage-gated sodium channels^{14,15}. We achieved this by the manufacture of two poly-cistronic constructs, allowing the independent expression of *SCN9A* (wild-type or mutant) and a red fluorescent protein, in addition to *SCN1B*, *SCN2B* and a green fluorescent protein (Fig. 4a). HEK293 cells were transiently transfected with both constructs using lipofection, and cells exhibiting

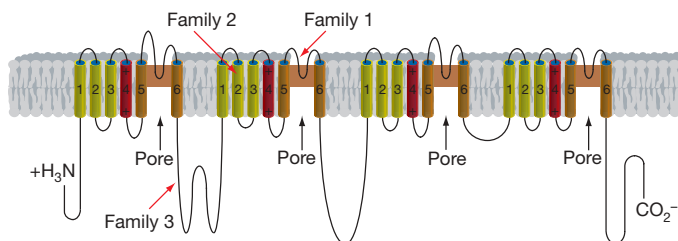


Figure 3 | Schematic representation of $\text{Na}_v1.7$, the voltage-gated sodium channel α -subunit encoded by *SCN9A*, and the locations of the identified human mutations. *SCN9A* encodes a plasma membrane protein: in the figure, the plasma membrane is shown in grey; the extracellular region is uppermost; and intracellular region below. $\text{Na}_v1.7$ is predicted to fold into four similar domains with each domain comprising six α -helical transmembrane segments (labelled 1–6). Transmembrane segments 5 and 6 are the pore-lining segments and the voltage sensor is located in transmembrane segment 4 of each domain (depicted by a plus symbol). The red arrows indicate the location of the nonsense mutation in each family.

both red and green fluorescence were selected for electrophysiological studies in the expectation that such cells would also express *SCN9A*, *SCN1B* and *SCN2B* (Fig. 4b). Control cells were transfected with the $\beta_1\beta_2$ subunit construct alone and selected on the basis of

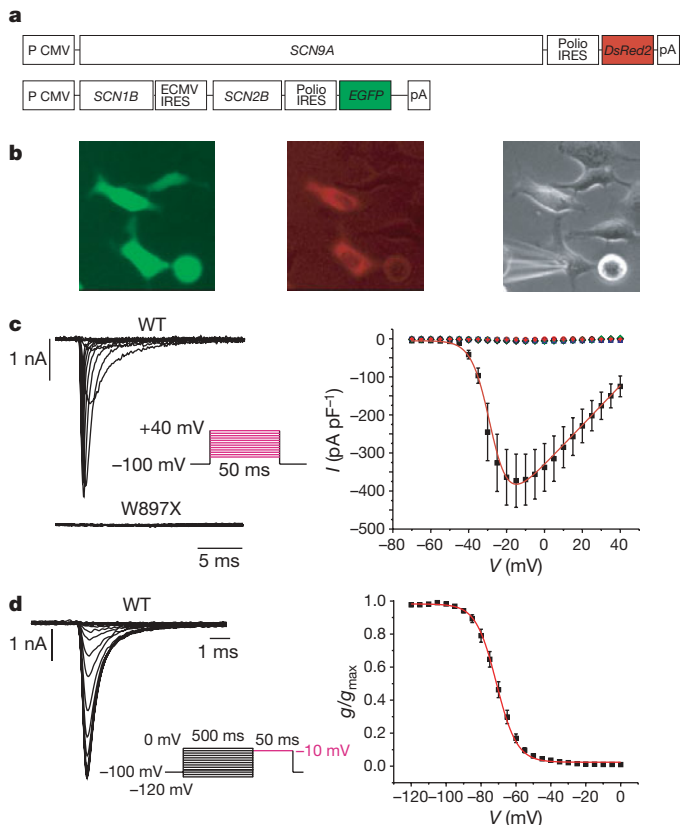


Figure 4 | Patch-clamping experiments to investigate the voltage-gated sodium channel activity of wild-type and truncated $\text{Na}_v1.7$. **a**, Constructs used in the patch-clamping experiments (see Supplementary Methods). We cloned wild-type *SCN9A* and then used site-directed mutagenesis to manufacture three further constructs each containing a family mutation. Each *SCN9A* construct was sequentially co-transfected with a plasmid containing the auxiliary sodium channel β_1 and β_2 subunits, and only cells clearly expressing DsRed2 (red fluorescence) and EGFP (green fluorescence) were measured for electrical activity. ECMV, encephalomyocarditis virus; IRES, internal ribosome entry site. **b**, A typical cell used in the patch-clamping experiments, showing both EGFP fluorescence (left) and DsRed2 fluorescence (middle), with the pipette attached in phase contrast (right). **c**, Left panel: initial current responses to 50-ms voltage steps of 5-mV increments between -70 and $+40$ mV from a holding potential of -100 mV, in a whole-cell voltage clamp recording applied at ~ 0.5 Hz for a cell co-expressing wild-type (WT) $\text{Na}_v1.7$ (top) or $\text{Na}_v1.7$ W897X (bottom) with the β -subunits. The inset shows the voltage pulse protocol. Right panel: current-voltage relationship of the peak currents normalized for cell size (pA per pF) obtained using the experimental set-up shown on the left. Black squares, wild type ($n = 13$); red circles, I767X ($n = 7$); blue squares, W897X ($n = 7$); green diamonds, S459X ($n = 5$); white diamonds, β -subunits only ($n = 5$). The red line represents a fit of the wild-type data with a Boltzmann equation $y = (A_2 + (A_1 - A_2)/(1 + \exp((V_{0.5} - x)/k)))(x - V_{\text{rev}})$, where $V_{0.5} = 28.0$ mV, $k = 4.9$ mV, $V_{\text{rev}} = 64$ mV. **d**, Left panel: voltage dependence of the steady-state inactivation of wild-type $\text{Na}_v1.7$ plus β -subunits was measured by holding the membrane potential for 500 ms at conditioning voltages from -120 to 0 mV (at 5-mV increments) before stepping to a test pulse at -10 mV for 50 ms. The inset shows the voltage pulse protocol, which was applied at 0.5 Hz. Only the current responses to the test pulse are shown. Right panel: the peak currents obtained as on the left were normalized to the maximum peak current, and plotted against the holding potential applied during the conditioning pulse. The red line represents a fit of the data with a Boltzmann equation $y = (A_1 - A_2)/(1 + \exp((x - V_{0.5})/k)) + A_2$, where $V_{0.5} = -71$ mV, $k = 5.9$ mV, $n = 13$. Error bars in **c** and **d** represent standard errors.

their green fluorescence. Whole-cell voltage clamp recordings from cells co-expressing wild-type $\text{Na}_v1.7$ with the $\beta_1\beta_2$ subunits, revealed a voltage-gated Na^+ current with a peak amplitude of $-373 \pm 70 \text{ pA pF}^{-1}$ at -15 mV ($n = 13$), compared with a background current of $-3 \pm 1 \text{ pA pF}^{-1}$ ($n = 5$) in cells transfected with the $\beta_1\beta_2$ construct alone ($P = 0.005$ for wild-type $\text{Na}_v1.7$ versus control) (Fig. 4c). The voltage dependence of activation of $\text{Na}_v1.7$ could be described by a Boltzmann function, with half-maximum activation ($V_{0.5}$) of $-28 \pm 1 \text{ mV}$, $k = 4.9 \pm 0.5 \text{ mV}$ and a reversal potential (V_{rev}) of $+64 \pm 3 \text{ mV}$ ($n = 13$). Voltage-dependent inactivation could also be described by a Boltzmann function, with half-maximal inactivation at $-71 \pm 1 \text{ mV}$ and $k = 5.9 \pm 0.4 \text{ mV}$ ($n = 13$) (Fig. 4d). These properties, as well as the voltage dependence of the kinetics of activation and inactivation (Supplementary Fig. 1), are similar to those described previously for $\text{Na}_v1.7$ currents^{6,16–18}. In contrast, cells co-transfected with each of the mutated $\text{Na}_v1.7$ subunits plus $\beta_1\beta_2$ exhibited currents that were not significantly different from those recorded from control cells (Fig. 4c). The mean peak currents at -15 mV were: I767X, $-6 \pm 2 \text{ pA pF}^{-1}$ ($n = 7$, $P > 0.1$ versus control); W897X, $-5 \pm 1 \text{ pA pF}^{-1}$ ($n = 7$, $P > 0.2$ versus control); S459X, $-5 \pm 2 \text{ pA pF}^{-1}$ ($n = 5$, $P > 0.2$ versus control). These results indicate that the absence of pain in these patients is co-incident with a complete loss of function of $\text{Na}_v1.7$. Given the proposed role of $\text{Na}_v1.7$ in action potential generation in DRG neurons^{11,19,20} and the lack of pain perception in these patients, our data suggest that the firing of action potentials may be substantially compromised in nociceptive neurons lacking $\text{Na}_v1.7$. Although we favour a defect in peripheral nociceptive transmission as the most likely explanation for the lack of pain perception in these patients, we cannot rule out the possibility that the phenotype could be related to a $\text{Na}_v1.7$ -mediated central nervous system defect. *SCN9A* is known to be expressed in human⁸ and monkey²¹ spinal cord and brain, albeit at a much lower level than in dorsal root ganglia²¹.

Conclusions and prospects

SCN9A has previously been shown to be involved in nociception in both humans and rodents. The autosomal-dominant pain disorder primary erythralgia (OMIM 133020), characterized by severe, episodic burning pain in the extremities in response to warm stimuli or moderate exercise, is caused by gain of function mutations in *SCN9A*²². These mutations alter the threshold of activation of the $\text{Na}_v1.7$ sodium channel, resulting in hyperexcitability of pain signalling neurons^{16,23,24}. Loss of function mutations in $\text{Na}_v1.7$ have not been reported previously in humans, but have been studied in mice that lack $\text{Na}_v1.7$ in nociceptive neurons^{25,26}. Such mice show increased mechanical and thermal pain thresholds and striking deficits in the development of inflammatory pain symptoms. Global $\text{Na}_v1.7$ -null mutant mice, however, die shortly after birth²⁵, associated with a failure to feed, in stark contrast to the human families where there is no reported increase in early mortality. The amino acid sequences of human and mouse $\text{Na}_v1.7$ proteins are highly conserved (95% similarity and 92% identity, Supplementary Fig. 2), as is the domain structure. Why the global- $\text{Na}_v1.7$ -deficient mice should die when the humans do not is therefore unclear. It might, however, reflect species differences in early postnatal development or in the expression pattern or function of $\text{Na}_v1.7$ in non-nociceptive neurons.

Pain perception has evolved as a protective mechanism that allows an organism to detect tissue damage and then modulate its activity while repair occurs. This is illustrated by the nociceptive neuropathies 'congenital insensitivity to pain with anhidrosis (HSAN4)' (OMIM 256800) and 'congenital insensitivity to pain (HSAN5)', where affected individuals often suffer permanent injury during childhood because they fail to notice illnesses or injuries, and fail to learn pain-avoiding (severe risk-avoiding) behaviours—as in this study's index case. Because pain perception pathways are so numerous and complex it was surprising that disruption of a single gene, *SCN9A*, would lead to a complete loss of nociceptive input. We show

here that this is the case, at least in humans. Given the key role of *SCN9A* in human pain perception, it is interesting to speculate whether single-nucleotide polymorphisms in *SCN9A* may explain variations in pain thresholds between individuals. Finally, as individuals with null mutations in *SCN9A* are otherwise healthy, drugs blocking $\text{Na}_v1.7$ function have the potential to produce new and potentially safer analgesia^{27,28}.

METHODS

Clinical note. We had originally diagnosed the study families as having congenital indifference to pain, but we subsequently found difficulty in the nomenclature of nociceptive disorders.

There are two schools of thought in the literature for distinguishing between congenital insensitivity to pain and congenital indifference to pain. On one side the distinction between insensitivity and indifference to pain is centred on the respective presence or absence of a neuropathy, this being the cardinal feature in insensitivity^{2,5}. The families we studied were diagnosed as having indifference owing to an absence of neuropathy in our patients. Also, the phenotype in our patients closely resembles that seen in other patients described with indifference: 'not experiencing pain anywhere over their bodies from needle-prick or injury, but could recognize the point from the head of a pin, had no physiological responses to noxious stimuli, no evidence of neurological abnormality and biopsy and post-mortem examination did not disclose a structural abnormality of nerve, spinal cord or brain'²².

On the other side the distinction outlines that indifference implies a lack of concern to a stimulus that is received and perceived (the sensory pathways are considered normal), whereas insensitivity describes the absence of painful sensation or failure to receive perception due to a detectable defect of sensory pathways³. From this point of view, the families in this study clearly have insensitivity, as they have no personal knowledge of pain.

The distinction between insensitivity and indifference can therefore be made on either histopathological grounds of finding a neuropathy, or on clinical grounds of whether the pain is felt and ignored or not felt at all; these approaches yield different diagnostic designations for our families.

The families reported here have no nociception (the ability to respond to tissue damage) and *SCN9A* null mutations. The problem is that loss of function of $\text{Na}_v1.7$ is a molecular defect that is not routinely detected by histopathology; that is, the patients have a normal nerve biopsy and so in the clinic they would probably be diagnosed as presenting with indifference rather than insensitivity to pain. However, the patients do not seem to fit the indifference diagnostic criteria of 'detecting pain but not reacting to it adversely'. We therefore propose to call this disorder channelopathy-associated insensitivity to pain, which encompasses the underlying disease pathogenesis (a channelopathy affecting the nociceptive system) and the primary clinical feature (an inability to perceive nociceptive pain).

Microsatellite mapping and linkage analysis. Initial autozygosity mapping was performed using DNA samples from the affected individuals and their parents for families 1 and 2, and a panel of 400 polymorphic microsatellite markers, originally derived from the Weber Mapping Panel 6 and subsequently modified in-house. Family 3 was later similarly studied. The genotype data was analysed by hand for inconsistencies and by family haplotype to identify regions of homozygosity. To calculate a LOD score, the disease was analysed as an autosomal-recessive trait with a mutant allele frequency of 0.001, equal sex recombination frequencies and allele frequencies of $1/n$ (n = number of different alleles observed). Two-point LOD scores were calculated using MLINK from FASTLINK 5.1.

Novel polymorphic markers. All novel polymorphic markers were identified in genomic sequence by use of Tandem Repeats Finder and the UCSC Human Genome Browser. Primers were designed using Primer3 and checked for specificity by BLAST.

Mutation detection. All homozygous mutations were initially detected by bi-directional sequencing of genomic DNA using standard methods. Sequencing primers were designed using Primer3, BLAST and the sequence of AC107082 and AC108146. The mutations were confirmed to segregate in the families and to be absent from 150 ethnically matched human controls as follows: the family 2 (2298delT) and the family 3 (1376C→G) mutations were bi-directionally sequenced; amplification refractory mutation system (ARMS) polymerase chain reaction of the wild-type and mutant sequences was performed for the family 1 (2691G→A) mutation using the wild-type primer pair (5' to 3') TGCTTTACCTTTTGAACAAAA and TGGAGAAGTCGTTTCATCTGC (product = 314 bp) and the mutant primer pair CTGTACGCTCCACGGAGA and CATCACAATAATTTCACAGAGA (product = 224 bp).

Electrophysiology. HEK293A cells (QBiogene), cultured in DMEM supplemented with 5% FCS, were transiently transfected with plasmids expressing either wild-type or mutant *SCN9A* plus *DsRed2* and/or *SCN1B* plus *SCN2B* plus *EGFP* using lipofectamine 2000 (see Supplementary Methods for cloning). Experiments were performed 2–3 days after transfection on cells positive for *DsRed2* and *EGFP* (or *EGFP* for β -subunit only control). Fluorescence was detected with excitation at 550 ± 7 nm and 488 ± 5 nm, respectively, using appropriate emission filters and MetaMorph (Molecular Devices) software controlling a monochromator (Cairn) and a CCD-camera (Orca ER, Hammamatsu) mounted on an Olympus IX71 microscope with a $\times 40$ objective. Micro-electrodes were pulled from borosilicate glass (GC150T, Harvard Apparatus) and the tips coated with melted beeswax. Electrodes were fire-polished using a microforge (Narishige) and had resistances of 2.5–3 M Ω when filled with pipette solution. Standard whole-cell currents were filtered at 10 kHz and recorded at 20 kHz at 22–24 °C using an EPC10 amplifier controlled by patchmaster software (HEKA Electronic). The holding potential was -100 mV, 70% series resistance compensation was used throughout, and currents were zero- and leak-subtracted using a p/4 protocol. Analysis was performed with pulsefit (HEKA Electronic) and Origin software (OriginLab Corp.). The bath solution contained (in mM): 3 KCl, 140 NaCl, 2 CaCl₂, 1 MgCl₂, 10 HEPES, 1 glucose (pH 7.4 with NaOH). The patch pipette solution contained (in mM): 107 CsF, 10 NaCl, 1 CaCl₂, 2 MgCl₂, 10 HEPES, 10 TEACl, 10 EGTA (pH 7.2 with CsOH).

Ethical and licensing considerations. The Cambridge Research Ethics Committee approved the study. There are no licensing considerations.

List of URLs. The UCSC Human Genome Browser is available at <http://genome.ucsc.edu/cgi-bin/hgGateway>; Tandem Repeats Finder is available at <http://tandem.bu.edu/trf/trf.submit.options.html>; Primer3 is available at http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi; LALIGN is available at http://www.ch.embnet.org/software/LALIGN_form.html; InterProScan is available at <http://www.ebi.ac.uk/InterProScan>; FASTLINK 5.1 is available at <http://linkage.rockefeller.edu/soft/fastlink>; BLAST is available at <http://www.ncbi.nlm.nih.gov/BLAST>.

Received 9 August; accepted 3 November 2006.

1. Dearborn, G. A case of congenital general pure analgesia. *J. Nerv. Ment. Dis.* **75**, 612–615 (1932).
2. Dyck, P. J. *et al.* Not 'indifference to pain' but varieties of hereditary sensory and autonomic neuropathy. *Brain* **106**, 373–390 (1983).
3. Landrieu, P., Said, G. & Allaire, C. Dominantly transmitted congenital indifference to pain. *Ann. Neurol.* **27**, 574–578 (1990).
4. Nagasako, E. M., Oaklander, A. L. & Dworkin, R. H. Congenital insensitivity to pain: an update. *Pain* **101**, 213–219 (2003).
5. Klein, C. J., Sinnreich, M. & Dyck, P. J. Indifference rather than insensitivity to pain. *Ann. Neurol.* **53**, 417–418 (2003); author reply *Ann. Neurol.* **53**, 418–419 (2003).
6. Klugbauer, N., Lacinova, L., Flockerzi, V. & Hofmann, F. Structure and functional expression of a new member of the tetrodotoxin-sensitive voltage-activated sodium channel family from human neuroendocrine cells. *EMBO J.* **14**, 1084–1090 (1995).
7. Toledo-Aral, J. J. *et al.* Identification of PN1, a predominant voltage-dependent sodium channel expressed principally in peripheral neurons. *Proc. Natl Acad. Sci. USA* **94**, 1527–1532 (1997).
8. Sangameswaran, L. *et al.* A novel tetrodotoxin-sensitive, voltage-gated sodium channel expressed in rat and human dorsal root ganglia. *J. Biol. Chem.* **272**, 14805–14809 (1997).
9. Rush, A. M. *et al.* A single sodium channel mutation produces hyper- or hypoexcitability in different types of neurons. *Proc. Natl Acad. Sci. USA* **103**, 8245–8250 (2006).
10. Amir, R. *et al.* The role of sodium channels in chronic inflammatory and neuropathic pain. *J. Pain* **7**, S1–S29 (2006).

11. Cummins, T. R., Howe, J. R. & Waxman, S. G. Slow closed-state inactivation: a novel mechanism underlying ramp currents in cells expressing the hNE/PN1 sodium channel. *J. Neurosci.* **18**, 9607–9619 (1998).
12. Amrani, N., Sachs, M. S. & Jacobson, A. Early nonsense: mRNA decay solves a translational problem. *Nature Rev. Mol. Cell Biol.* **7**, 415–425 (2006).
13. Black, J. A. *et al.* Spinal sensory neurons express multiple sodium channel α -subunit mRNAs. *Brain Res. Mol. Brain Res.* **43**, 117–131 (1996).
14. Catterall, W. A. From ionic currents to molecular mechanisms: the structure and function of voltage-gated sodium channels. *Neuron* **26**, 13–25 (2000).
15. Isom, L. L. Sodium channel β subunits: anything but auxiliary. *Neuroscientist* **7**, 42–54 (2001).
16. Cummins, T. R., Dib-Hajj, S. D. & Waxman, S. G. Electrophysiological properties of mutant Nav1.7 sodium channels in a painful inherited neuropathy. *J. Neurosci.* **24**, 8232–8236 (2004).
17. Catterall, W. A., Goldin, A. L. & Waxman, S. G. International Union of Pharmacology. XLVII. Nomenclature and structure-function relationships of voltage-gated sodium channels. *Pharmacol. Rev.* **57**, 397–409 (2005).
18. Herzog, R. I., Cummins, T. R., Ghassemi, F., Dib-Hajj, S. D. & Waxman, S. G. Distinct repriming and closed-state inactivation kinetics of Nav1.6 and Nav1.7 sodium channels in mouse spinal sensory neurons. *J. Physiol. (Lond.)* **551**, 741–750 (2003).
19. Renganathan, M., Cummins, T. R. & Waxman, S. G. Contribution of Na(v)1.8 sodium channels to action potential electrogenesis in DRG neurons. *J. Neurophysiol.* **86**, 629–640 (2001).
20. Blair, N. T. & Bean, B. P. Roles of tetrodotoxin (TTX)-sensitive Na⁺ current, TTX-resistant Na⁺ current, and Ca²⁺ current in the action potentials of nociceptive sensory neurons. *J. Neurosci.* **22**, 10277–10290 (2002).
21. Raymond, C. K. *et al.* Expression of alternatively spliced sodium channel α -subunit genes. Unique splicing patterns are observed in dorsal root ganglia. *J. Biol. Chem.* **279**, 46234–46241 (2004).
22. Yang, Y. *et al.* Mutations in *SCN9A*, encoding a sodium channel α subunit, in patients with primary erythralgia. *J. Med. Genet.* **41**, 171–174 (2004).
23. Dib-Hajj, S. D. *et al.* Gain-of-function mutation in Nav1.7 in familial erythromelalgia induces bursting of sensory neurons. *Brain* **128**, 1847–1854 (2005).
24. Han, C. *et al.* Sporadic onset of erythralgia: a gain-of-function mutation in Nav1.7. *Ann. Neurol.* **59**, 553–558 (2006).
25. Nassar, M. A. *et al.* Nociceptor-specific gene deletion reveals a major role for Nav1.7 (PN1) in acute and inflammatory pain. *Proc. Natl Acad. Sci. USA* **101**, 12706–12711 (2004).
26. Nassar, M. A., Levato, A., Stirling, L. C. & Wood, J. N. Neuropathic pain develops normally in mice lacking both Nav1.7 and Nav1.8. *Mol. Pain* **1**, 24 (2005).
27. Akopian, A. N., Abson, N. C. & Wood, J. N. Molecular genetic approaches to nociceptor development and function. *Trends Neurosci.* **19**, 240–246 (1996).
28. Scholz, J. & Woolf, C. J. Can we conquer pain? *Nature Neurosci.* **5** (Suppl), 1062–1067 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the families who participated in this study, A. Boylston for critical advice, and Pfizer, the Wellcome Trust and St John's College, Cambridge for funding.

Author Contributions This study was designed by J.J.C., F.R., E.R., R.W., D.P.M., F.M.G. and C.G.W.; patient identification and phenotype assessment was performed by G.K., H.J., J.M., Y.R., L.A.-G., H.H., E.M.V., S.G. and C.G.W.; DNA extraction, linkage analysis, bioinformatics and sequencing was performed by J.J.C., A.K.N., E.R., K.S. and C.G.W.; cloning was performed by J.J.C.; electrophysiology was performed by F.R. and F.M.G.; and the paper was written by J.J.C., F.R., G.T., J.N.W., F.M.G. and C.G.W.

Author Information The sequence for full-length human *SCN9A* cloned from fetal brain mRNA is deposited in GenBank under accession number DQ857292. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.G.W. (cw347@cam.ac.uk).

Regulation of the bacterial cell cycle by an integrated genetic circuit

Emanuele G. Biondi^{1†}, Sarah J. Reisinger², Jeffrey M. Skerker^{1†}, Muhammad Arif¹, Barrett S. Perchuk^{1†}, Kathleen R. Ryan² & Michael T. Laub^{1†}

How bacteria regulate cell cycle progression at a molecular level is a fundamental but poorly understood problem. In *Caulobacter crescentus*, two-component signal transduction proteins are crucial for cell cycle regulation, but the connectivity of regulators involved has remained elusive and key factors are unidentified. Here we identify ChpT, an essential histidine phosphotransferase that controls the activity of CtrA, the master cell cycle regulator. We show that the essential histidine kinase CckA initiates two phosphorelays, each requiring ChpT, which lead to the phosphorylation and stabilization of CtrA. Downregulation of CckA activity therefore results in the dephosphorylation and degradation of CtrA, which in turn allow the initiation of DNA replication. Furthermore, we show that CtrA triggers its own destruction by promoting cell division and inducing synthesis of the essential regulator DivK, which feeds back to downregulate CckA immediately before S phase. Our results define a single integrated circuit whose components and connectivity can account for the cell cycle oscillations of CtrA in *Caulobacter*.

Although the cell cycle of eukaryotes has been described in molecular detail, the bacterial cell cycle remains poorly understood. Genome sequencing projects have shown that the main regulators of the eukaryotic cell cycle, such as cyclin-dependent kinases, are not found in bacteria. How then do bacteria regulate cell cycle progression, and is the logic of the underlying regulatory circuit similar despite the use of different molecules?

The bacterium *Caulobacter crescentus* is an attractive model for examining cell cycle regulation in bacteria^{1,2}. *Caulobacter* is easily synchronized and follows a pattern of once-and-only-once replication so that the G1, S and G2 phases are temporally distinguished, as in eukaryotes. Moreover, each cell division in *Caulobacter* is asymmetric and produces daughter cells—stalked and swarmer cells—that are committed to different stages of the cell cycle. Cell cycle regulation in *Caulobacter* relies on two-component signal transduction systems, comprising histidine kinases and their response regulator substrates³. Many two-component signalling genes have been identified in genetic screens for cell cycle regulators^{2,4}. However, the precise biochemical connectivity of these proteins is unknown.

The master regulator of the *Caulobacter* cell cycle is CtrA, an essential response regulator whose activity as a transcription factor varies as a function of the cell cycle^{5,6}. In G1 cells, CtrA is present and phosphorylated (CtrA~P), enabling it to bind and silence the origin of replication⁷. At the G1–S transition, CtrA~P is dephosphorylated and degraded, thereby freeing the origin and permitting the initiation of DNA replication⁵. Clearing CtrA from the cell also leads to the synthesis of GcrA, which accumulates and triggers *de novo* transcription of *ctrA*^{8,9}. The initial synthesis of CtrA leads to positive transcriptional autoregulation and a burst of CtrA synthesis in late stalked cells¹⁰. The newly synthesized CtrA is phosphorylated and is not immediately subject to proteolysis. In predivisional cells, CtrA~P drives the expression of more than 50 genes, many of which are required for completing the cell cycle^{9,11}.

The reciprocal regulation of CtrA and GcrA at a transcriptional level has been suggested to form an oscillating genetic circuit that

drives the cell cycle^{8,12}. However, cells that constitutively transcribe *ctrA* still proceed through the cell cycle, indicating that transcriptional control of *ctrA* is not strictly necessary⁵. Instead, regulation of CtrA activity by either temporally controlled phosphorylation or proteolysis is required to drive cell cycle progression⁵. Both proteolysis and dephosphorylation of CtrA occur at the G1–S transition and in the stalked compartment of the predivisional cell, just before cell separation^{5,13}. Cells that synthesize CtrA(D51E)Δ3Ω, a version of CtrA that cannot be proteolysed and that mimics the phosphorylated state, arrest in G1 because CtrA activity must be temporarily eliminated to allow DNA replication. The histidine kinase CckA is required *in vivo* for the phosphorylation of CtrA¹⁴, but a direct biochemical link between the two has not been shown. Moreover, the autophosphorylation of CckA is cell cycle-regulated and correlates with the protein's polar localization¹⁵, but factors that influence CckA activity have not been identified.

Another essential cell cycle regulator in *Caulobacter* is the single-domain response regulator DivK¹⁶. The cold-sensitive mutant *divK341cs* (*divK*^{cs}) arrests in G1 when incubated at the restrictive temperature. The similarity between this phenotype and that of cells that synthesize CtrA(D51E)Δ3Ω indicates that DivK might somehow promote both the proteolysis and dephosphorylation of CtrA in preparation for DNA replication. However, no model has emerged to suggest how DivK regulates both proteolysis and dephosphorylation of CtrA.

DivK has also been implicated in controlling cellular asymmetry and differentiation. DivJ, the primary kinase for DivK, is located at the stalked pole whereas PleC, a phosphatase for DivK~P, resides at the swarmer pole^{16–18}. Septum formation and cell division therefore produce daughter cells with different levels of DivK~P, with higher levels in the stalked cell than in the swarmer cell¹⁹. But it is unclear how this difference in DivK~P levels in the daughter cells is translated into a difference in CtrA activity and in cell cycle position.

Here, we use a combination of biochemical, genetic and cell biological assays to map the connectivity of a regulatory network

¹FAS Center for Systems Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ²Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, California 94720, USA. †Present address: Massachusetts Institute of Technology, Department of Biology, Cambridge, Massachusetts 02139, USA.

comprising two-component signalling proteins that can account for cell cycle oscillations in *Caulobacter*.

ChpT mediates a phosphorelay from CckA to CtrA

Hybrid kinases, such as CckA, can directly phosphorylate a response regulator or initiate a phosphorelay in which a phosphoryl group is passed intramolecularly to the kinase's own receiver domain, then to a histidine phosphotransferase (HPT), and finally to a soluble response regulator. To test whether CckA phosphorylates CtrA directly, we used phosphotransfer profiling²⁰ in which a purified kinase domain is tested, in parallel, for phosphotransfer to each of the 44 purified response regulators from *Caulobacter*. As histidine kinases exhibit a kinetic preference *in vitro* for their *in vivo* cognate substrates, this technique allows the rapid identification of targets for a given kinase²⁰. In a 5-min incubation, the kinase domain of CckA (CckA-HK) did not transfer phosphate to CtrA, but it did transfer phosphate to four other response regulators (Fig. 1a), including its own receiver domain (CckA-RD). At earlier time points CckA-HK transferred only to PetR (CC2931) and CckA-RD (see Supplementary Fig. 1a). We conclude that CtrA is not phosphorylated directly by CckA-HK, and that a phosphorelay probably leads from CckA to CtrA through an unknown HPT. Whether PetR is a target of CckA *in vivo* remains to be shown.

HPTs are difficult to identify by sequence homology as they require conservation of only a small number of crucial residues, and none were predicted in the original annotation of the *Caulobacter* genome²¹. To identify putative HPTs for the CtrA pathway, we searched the *Caulobacter* genome for predicted proteins that shared the following characteristics of known HPTs: (1) <250 amino acids in length; (2) >70% α -helical; and (3) a histidine residue within a predicted α -helix. These criteria yielded more than 50 genes. We further predicted that an HPT that mediated a phosphorelay from CckA to CtrA should be present only in organisms that also contain orthologues of *cckA* and *ctrA*. Imposing this criterion yielded a single candidate, CC3470, which we name *chpT* for cell cycle histidine phosphotransferase (see Supplementary Fig. 2).

We predicted that an HPT that connected the essential signalling proteins CckA and CtrA would also be essential for viability. Consistent with this prediction, we obtained stable deletions of *chpT* only when a copy of the gene was provided *in trans* on a plasmid (see Supplementary Table 1). To determine the effects of *chpT* depletion, we constructed a strain in which the only copy of *chpT* is under the control of P_{xyb} , a xylose-inducible promoter²². This strain doubled every 150 min in rich medium containing xylose (Fig. 2a; wild-type doubling time was 90 min), and individual cells grown in xylose were elongated (Fig. 2b). These differences from the wild type are probably due to substitution of the native *chpT* promoter with P_{xyb} . When shifted to medium containing glucose, which represses P_{xyb} , the *chpT* depletion strain began to lose viability within 2–3 h (Fig. 2a), supporting the conclusion that *chpT* is an essential gene. After the shift, cells from the *chpT* depletion strain became filamentous, in a manner strikingly similar to *ctrA^{ts}* and *cckA^{ts}* strains grown at the restrictive temperature (Fig. 2b). Using whole-genome DNA microarrays, we found that the global pattern of gene expression in the *chpT* depletion strain was highly correlated with the expression patterns seen in *ctrA^{ts}* and *cckA^{ts}* strains (see Supplementary Table 2, Supplementary Fig. 3). Together, these data show that depletion of *chpT* produces a phenotype similar to *ctrA* and *cckA* mutants, supporting the conclusion that *chpT* encodes an HPT that lies between CckA and CtrA.

To confirm the existence of a CckA–ChpT–CtrA pathway, we measured levels of CtrA protein and phosphorylation *in vivo* in the *chpT* depletion strain. CtrA protein was present at roughly twofold lower levels in the *chpT* depletion strain than in the wild type (Fig. 2c). However, CtrA phosphorylation was even lower, dropping below the level of detection in the P_{xyb} –*chpT* strain (Fig. 2c).

Two CckA/ChpT-dependent phosphorelays

To prove that CckA–ChpT–CtrA comprises a phosphorelay, we examined phosphotransfer relationships among purified components of the pathway *in vitro*. Incubation of purified CckA-HK,

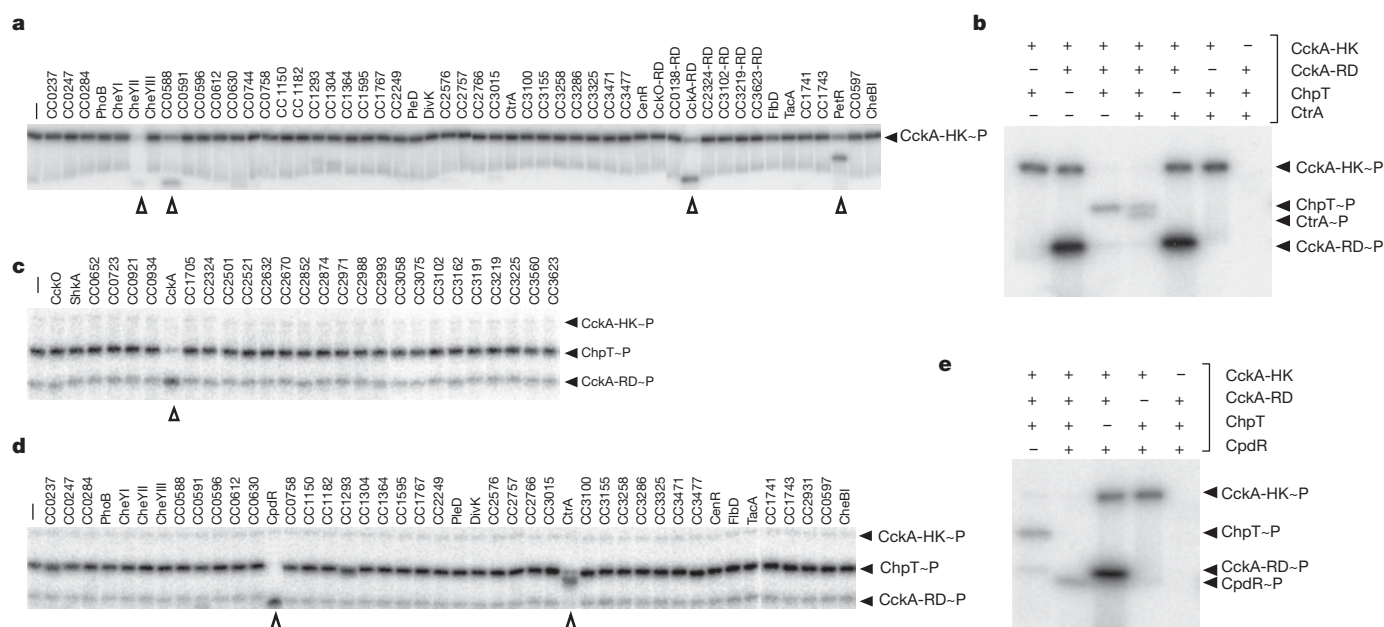


Figure 1 | Identification and *in vitro* reconstitution of two phosphorelays controlling CtrA. **a**, Phosphotransfer profiling²⁰ of CckA. The purified kinase domain of CckA (CckA-HK) was autophosphorylated and mixed with each of the 44 purified response regulators and with 7 of the 27 receiver domains from hybrid histidine kinases encoded in the *C. crescentus* genome. Open arrowheads indicate lanes with high-efficiency phosphotransfer, manifested as loss of radiolabel from the CckA-HK~P band and/or

incorporation of label on a response regulator or a receiver domain. **b**, Reconstitution of the CckA–ChpT–CtrA phosphorelay *in vitro*. Pluses and minuses indicate the presence or absence of reaction components. **c**, **d**, Phosphotransfer profiling of ChpT~P against **(c)** the 27 receiver domains of hybrid histidine kinases and **(d)** the 44 response regulators encoded in the *C. crescentus* genome. **e**, Reconstitution of the CckA–ChpT–CpdR phosphorelay *in vitro*.

CckA-RD, ChpT and CtrA with [γ - 32 P]ATP led to accumulation of radiolabel in CtrA (Fig. 1b). The order of phosphotransfer was determined by omitting individual components (Fig. 1b). These *in vitro* data confirm a phosphorelay in which CckA-HK autophosphorylates and then passes a phosphoryl group intramolecularly to CckA-RD, then to ChpT, and finally to CtrA.

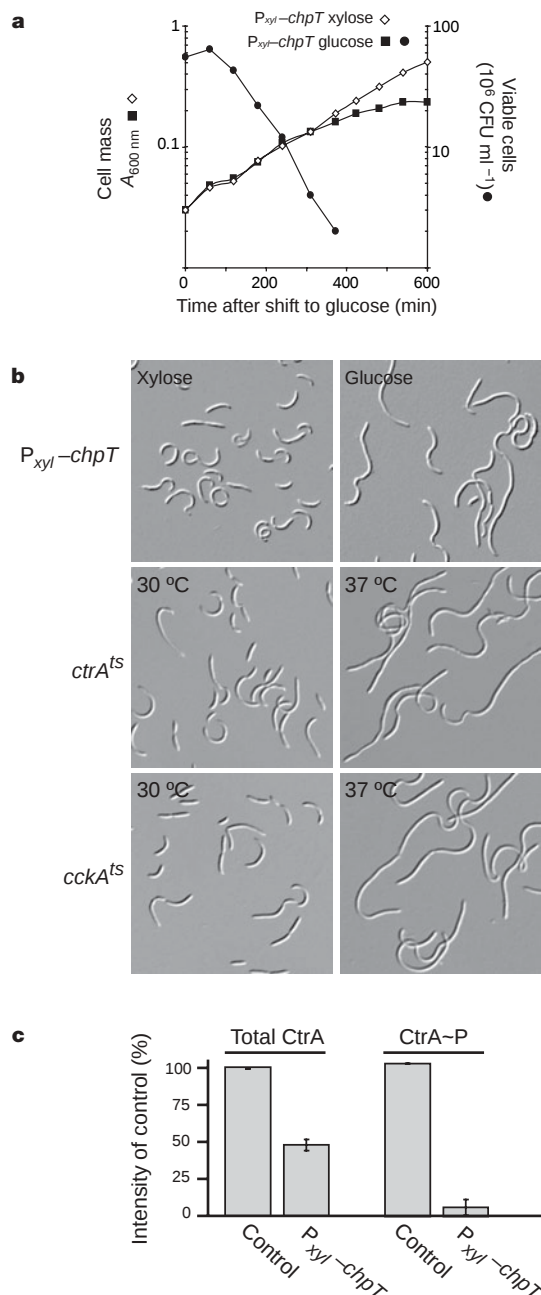


Figure 2 | *chpT* encodes a histidine phosphotransferase that is essential for viability and phenocopies *ctrA*^{ts} and *cckA*^{ts}. **a**, *chpT* depletion strain (ML808: P_{xyI} -*chpT*) was grown to mid-log phase in xylose and then shifted to glucose or maintained in xylose. Growth (absorbance, $A_{600\text{ nm}}$) and viability (colony forming units, CFU) were measured for up to 600 min. P_{xyI} -*chpT* cells grown in xylose accumulated CFUs at approximately the same rate as wild-type cells (data not shown). **b**, Morphology of *chpT* depletion strain grown for 6 h in xylose or glucose and compared to *ctrA*^{ts} (*ctrA*V148F) and *cckA*^{ts} (*cckA*TS1) grown at 30 °C and 37 °C for 6 h. **c**, Levels of total CtrA protein and CtrA~P in the *chpT* depletion strain, grown in xylose, relative to a control strain bearing an empty vector. Error bars represent the mean $\pm$ s.d. from three independent experiments.

To determine whether CckA and CtrA are the exclusive partners of ChpT, we used phosphotransfer profiling^{20,21}. We first examined the ability of ChpT~P to transfer phosphate to the purified receiver domains from each of the 27 hybrid histidine kinases encoded in the *Caulobacter* genome. ChpT had a single, highly preferred substrate, the CckA receiver domain (Fig. 1c), indicating that CckA is the only input to ChpT. Next, we profiled ChpT~P against each of the 44 purified, soluble *Caulobacter* response regulators. In this case ChpT~P efficiently phosphorylated two response regulators, CtrA and CpdR²³ (Fig. 1d). Detailed time courses showed that ChpT~P transferred phosphate to these regulators at approximately equal rates (see Supplementary Fig. 1b). These data confirmed the existence of a CckA–ChpT–CtrA phosphorelay and identified a second phosphorelay with shared components, leading from CckA to ChpT to CpdR (Fig. 1e).

DivK feeds back to control CckA

The CckA–ChpT–CtrA phosphorelay culminates in the phosphorylation of CtrA and its activation as a transcription factor. The CckA–ChpT–CpdR phosphorelay culminates in the phosphorylation of CpdR, which prevents CpdR from triggering the proteolysis of CtrA²³ (see Supplementary Information Note 1, Supplementary Fig. 4). Our results therefore show that two phosphorelays, both stemming from CckA and ChpT, simultaneously regulate the phosphorylation and proteolysis of CtrA. However, the essential response regulator DivK has also been implicated in controlling these two processes. A cold-sensitive, loss-of-function mutation in *divK* (*divK*^{cs}) causes cells to arrest in G1 at the restrictive temperature²⁴. This phenotype is also seen in cells expressing *ctrA*(D51E)/ Δ 3 Ω , the constitutively active allele of *ctrA*, which might indicate that DivK is normally required to trigger CtrA dephosphorylation and degradation at the G1–S transition^{16,17,24}. Given our evidence that these events are controlled simultaneously by CckA and ChpT, we hypothesized that DivK does not control CtrA directly, but rather inhibits the activity of CckA.

To test this hypothesis, we measured the level of CckA~P *in vivo* in wild-type and *divK*^{cs} cells. Relative to the wild type, *divK*^{cs} cells contained ~1.8-fold more CckA~P per cell at the permissive temperature, and at least fourfold more CckA~P at the restrictive temperature (Fig. 3a). This indicates that DivK is required to inactivate CckA at the G1–S transition.

Next, we examined the effect of DivK on the subcellular localization of CckA (see Supplementary Note 2). CckA is dynamically localized in a pattern that parallels its *in vivo* kinase activity, which might indicate that CckA is active when localized to polar regions^{14,15}. CckA is localized to the swarmer pole and is active in early G1 cells, is delocalized and downregulated before the G1–S transition, and then is bipolarly localized and phosphorylated in the predivisional cell^{14,15} (Fig. 4b). CckA is then delocalized first in the stalked progeny, which enters S phase first (see Supplementary Figs 5, 9). As the delocalization of CckA is correlated with the G1–S transition, we asked whether DivK was required for this delocalization and hence for the onset of S phase. In a mixed population of *divK*^{cs} cells at the permissive temperature (37 °C), the pattern of CckA–GFP localization was similar to that of wild-type cells (Fig. 3b, c, Supplementary Fig. 6). However, when shifted to the restrictive temperature (24 °C) to induce G1-arrest²⁴, the *divK*^{cs} culture showed a marked increase in cells with a single bright focus of CckA–GFP at the stalked pole (Fig. 3b, c). Using time-lapse microscopy to examine individual cells, we found that *divK*^{cs} cells maintained at 24 °C continued to grow, but never lost polar localization of CckA–GFP and never divided (see Supplementary Fig. 7). However, when shifted to 37 °C, most cells (>70%) showed delocalization of CckA–GFP within 20 min, and 100% of these then divided within 120 min (see Supplementary Fig. 7).

To ensure that the effect of *divK*^{cs} on CckA localization was not a non-specific consequence of the G1-arrested state, we examined CckA–GFP localization in a strain that overproduces

CtrA(D51E) $\Delta 3\Omega$, which induces a G1-arrest independent of CckA^{5,15}. Most cells overproducing CtrA(D51E) $\Delta 3\Omega$ showed delocalization of CckA from the stalked pole (Fig. 3b, c). The constitutive localization of CckA in *divK^{cs}* cells is therefore not a consequence of G1-arrest, but is specifically due to the loss of DivK function.

Phosphorylation of DivK increases during the G1–S transition¹⁹, and immediately after cell division DivK is phosphorylated at higher levels in the new stalked cell than in the new swarmer cell^{25,26}. This indicates that phosphorylated DivK (DivK~P) might trigger the delocalization or inactivation of CckA. PleC is a phosphatase for DivK, whereas DivJ is the primary kinase^{16–18}. Our model predicts that cells lacking PleC (*pleC::Tn5*), which contain elevated levels of DivK~P, should show reduced polar localization of CckA. Conversely, $\Delta divJ$ cells, which contain reduced levels of DivK~P, should have CckA localized mainly to a single pole. Analysis of CckA–GFP localization in *pleC::Tn5* and $\Delta divJ$ strains confirmed these predictions (Fig. 3b, c). We also showed that CtrA~P levels were elevated in a *divK^{cs}* strain, but significantly reduced in *pleC::Tn5* (Fig. 3d), whereas a recent study showed that CtrA~P levels increase in a $\Delta divJ$ mutant²⁷. These results support the conclusion that DivK~P causes the delocalization and inactivation of CckA and hence the inactivation of CtrA, which is required for the G1–S transition.

Cell division partitions DivJ and PleC into separate compartments—the stalked and swarmer cells, respectively^{18,25}. Accord-

ingly, cell division should be essential for DivK~P to trigger the delocalization of CckA in a new stalked cell. To test this hypothesis, we examined the effect of blocking cell division on the dynamics of CckA–GFP localization. When stalked cells were depleted of the essential cell division protein FtsZ²⁸, CckA–GFP remained localized at the pole for extended periods of time (see Supplementary Fig. 8). Restoring *ftsZ* expression after prolonged depletion restored cell division and delocalization of CckA–GFP (see Supplementary Fig. 9). These results support the conclusion that cell division is required to activate the DivK-dependent feedback inhibition of CckA and CtrA. As CtrA is necessary for cell division^{6,29}, it ultimately helps to trigger feedback inhibition of itself through DivK.

CtrA directly activates *divK* expression late in the cell cycle^{9,24}. DivK is present throughout the cell cycle but increases in abundance late in the cell cycle¹⁹, coincident with the peak in transcription. The late induction of *divK* by CtrA might help to ensure that DivK does not accumulate to high levels and inhibit CckA during early stages of the cell cycle. If so, constitutive expression of *divK* should disrupt normal cell cycle oscillations by causing the inappropriate downregulation of CtrA. To test this, we generated strains in which the only copy of *divK* is expressed from a low-copy plasmid under the control of either its native, CtrA-regulated promoter, *P_{divK}*, or a constitutive promoter, *P_{xyI}*. These strains produce approximately equal levels of DivK (see Supplementary Fig. 10a), so the only difference is the promoter driving *divK*. The strain bearing *P_{divK}-divK* was similar

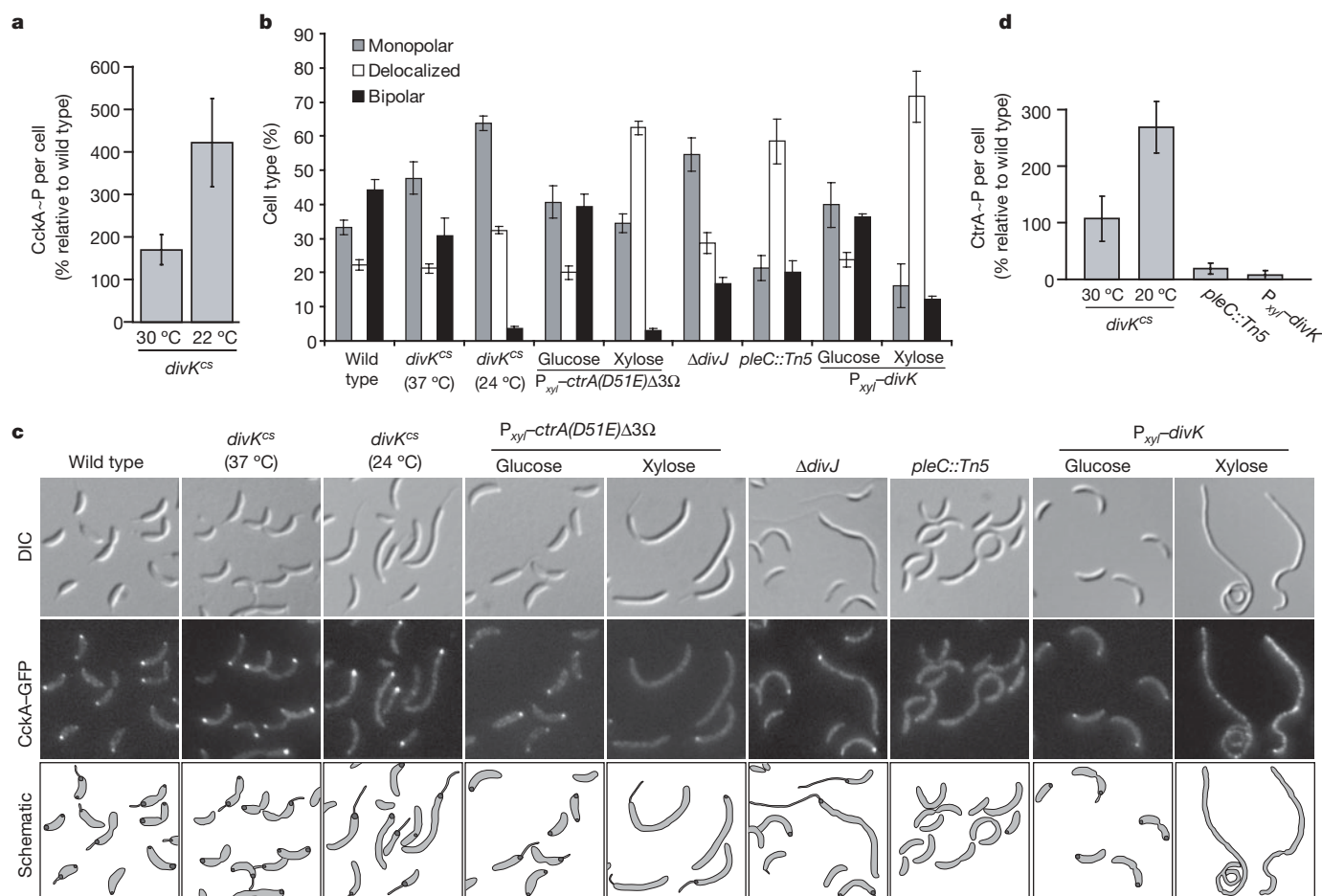


Figure 3 | DivK~P triggers the G1–S transition by causing inactivation and delocalization of CckA. **a**, Levels of CckA~P per cell in wild type and *divK^{cs}* cells. **b**, **c**, CckA–GFP localization was examined by epi-fluorescence microscopy in mixed populations of cells. For each strain, genotype and/or growth conditions are indicated in each panel. Localization patterns in a population of cells (see Supplementary Fig. 6) for each strain are quantified in **b** with representative images shown in **c**. Error bars represent the mean $\pm$

s.d. from three independent sets of 150 cells. **d**, Levels of CtrA~P per cell in *divK^{cs}* and *pleC::Tn5* strains relative to wild type (WT) and in cells overexpressing *divK* (WT + *pJS71-P_{xyI}-divK*) relative to cells bearing an empty vector (WT + *pJS71*). Error bars in **a** and **d** represent the mean $\pm$ s.d. from three independent experiments. DIC, differential interference contrast microscopy.

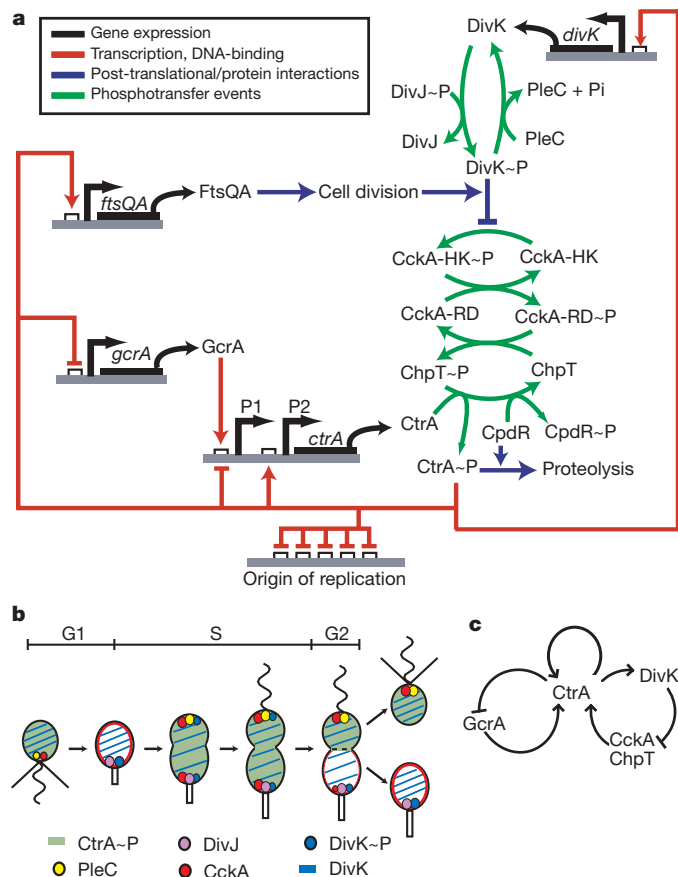


Figure 4 | Schematic of cell cycle regulation in *Caulobacter*. **a**, Diagram of the integrated genetic circuit controlling cell cycle progression and cellular asymmetry in *Caulobacter crescentus*. Biochemical relationships between components are colour-coded as indicated in the key. **b**, Summary of sub-cellular localization patterns for CckA, PleC, DivJ, CtrA~P, DivK and DivK~P during cell cycle progression. **c**, Summary of main feedback loops controlling CtrA activity and cell cycle progression.

to the wild type, whereas the strain bearing $P_{xyt}-divK$ had a severe cell cycle defect, similar to that seen in $ctrA^{ts}$ strains (see Supplementary Fig. 10b–d). These observations indicate that the CtrA-dependent transcription of *divK*, which occurs only late in the cell cycle, after CtrA has accumulated to high levels, helps to ensure the proper timing of DivK-mediated feedback on CtrA. We also found that expressing *divK* at higher levels completely disrupted cell cycle progression (see Supplementary Fig. 11a–d). Cells that overproduce DivK showed delocalization of CckA–GFP, a dramatic decrease in CtrA~P (Fig. 3b–d), and a terminal phenotype similar to $ctrA^{ts}$ strains. This phenotype depended on the presence of DivJ (see Supplementary Fig. 11b). These data further support the conclusion that DivK~P feeds back to downregulate CtrA.

Discussion

Our data establish an integrated molecular-level model of a regulatory network that accounts for *Caulobacter* cell cycle oscillations and the ability of a single cell to produce daughter cells committed to different cell cycle phases (Fig. 4). In stalked and predivisional cells, CckA is active and localized to both cell poles. Consequently, the two phosphorelays identified here (CckA–ChpT–CtrA and CckA–ChpT–CpdR) lead to CtrA phosphorylation and its protection from proteolysis. As CtrA~P accumulates, it triggers the expression of several genes that are required for late stages of the cell cycle, including *divK* and the essential cell division genes *ftsQ* and *ftsA*^{9,24,29}. The primary DivK kinase (DivJ) and phosphatase (PleC) are located at opposite ends of the predivisional cell such that daughter cells inherit one or

the other¹⁸. The stalked cell inherits DivJ, but not PleC, and can therefore accumulate phosphorylated DivK. This DivK~P triggers the delocalization and downregulation of CckA, thereby preventing the phosphorylation of CtrA and CpdR. Consequently, CtrA is dephosphorylated and degraded, permitting another round of DNA replication and expression of the CtrA-repressed gene *gcrA*. This eventually triggers new synthesis of CtrA and, hence, resets the cell cycle⁸. By contrast, the swarmer cell inherits PleC and dephosphorylates its pool of DivK. The lower levels of DivK~P allow CckA to remain localized and active, which stabilizes CtrA~P levels and blocks DNA replication initiation. As the swarmer cell develops, DivJ replaces PleC at the newly formed stalked pole^{18,30}, DivK~P accumulates, and CckA is delocalized and inactivated. As with the stalked cell, these steps allow the initiation of DNA replication, the expression of *gcrA*, and a resetting of the cell cycle. Although *cckA*, *chpT*, *ctrA* and *divK* are each essential for cell cycle progression, *divJ* and *pleC* mutants are viable, albeit with severe phenotypes (see Supplementary Fig. 12a–d), indicating that other factors might regulate DivK.

We propose that the integrated network identified here forms the basis of an oscillatory circuit that underlies cell cycle progression in *Caulobacter* (Fig. 4a). At the heart of this circuit is a negative feedback loop from CtrA to DivK and then back to CtrA, via CckA (Fig. 4c). However, negative feedback alone is insufficient to produce oscillations. Either nonlinearities or a time delay is also essential³¹. In our model, the accumulation of active DivK~P in the stalked progeny depends on cell division, introducing a time delay into the feedback loop. CtrA helps to activate cell division in predivisional cells²⁹, but cell division is delayed until after DNA replication and chromosome segregation have finished. The time needed to complete cell division thus enables CtrA to accumulate and persist at high levels in predivisional cells. Once cell division occurs, DivK~P can trigger the rapid turnover of CtrA and entry to S phase in the stalked cell. Inhibiting cell division was sufficient to prevent the downregulation of CckA and CtrA, indicating that cell division is the key time delay, but the temporal dynamics of the circuit must now be dissected in more detail.

Although many of the components of this cell cycle circuit were previously known, their connectivity had remained elusive. No single, coherent model had been produced to explain the rise and fall of CtrA activity, only its changes in transcription^{8,32}. Mapping the connectivity of the *Caulobacter* cell cycle regulatory network now allows us to compare it to that of eukaryotes at the level of regulatory architecture. For instance, the core oscillating machinery in *Caulobacter* involves a master regulator, CtrA, whose activity accumulates during the cell cycle. Our data show that CtrA triggers its own destruction by inducing *divK* transcription and cell division, which ultimately enable DivK~P to feedback and inhibit CckA and CtrA. In eukaryotes, an analogous regulatory strategy is employed where the activity of cyclin-dependent kinase accumulates during the cell cycle, ultimately triggering its own destruction³³. Recent work has also shown that two separable but interlinked oscillating circuits control yeast cell cycle progression³⁴. The multiple feedback loops present in the *Caulobacter* cell cycle circuit indicate that bacteria might also have evolved separate but interlaced oscillators (Fig. 4c). Such similarity in phylogenetically distant organisms could indicate that this regulatory architecture confers a strong selective advantage, regardless of the molecules used to implement it.

METHODS

Protein purifications, *in vitro* phosphorylation experiments, and phosphotransfer profiling were performed as described²⁰. For phosphorelay reconstitutions (Fig. 1b, e), reactions were initiated by the addition of [γ -³²P]ATP, incubated for 5 min, and analysed by SDS–PAGE (SDS–polyacrylamide gel electrophoresis) and phosphorimaging. Measurements of CtrA and CckA phosphorylation *in vivo* were performed as described^{5,15}. Briefly, equal numbers of cells were labelled for 10 min with [³²P]H₃PO₄, lysed, and immunoprecipitated with anti-CtrA or anti-CckA serum. Samples were analysed by SDS–PAGE and band intensities

quantified by phosphorimaging. Values per cell were obtained by normalizing to relative cell size. For measurements of CtrA and CtrA~P levels in the *chpT* depletion strain, cells were grown in the presence of xylose which produces a mild cell cycle phenotype (Fig. 2, Supplementary Fig. 3), but avoids the potential confounding effects of cell death during growth in glucose (Fig. 2a). For detailed methods, including growth conditions and strain construction, see Supplementary Information, Note 3.

Received 19 June; accepted 5 October 2006.

Published online 29 November 2006.

- McAdams, H. H. & Shapiro, L. A bacterial cell-cycle regulatory network operating in time and space. *Science* **301**, 1874–1877 (2003).
- Skerker, J. M. & Laub, M. T. Cell-cycle progression and the generation of asymmetry in *Caulobacter crescentus*. *Nature Rev. Microbiol.* **2**, 325–337 (2004).
- Stock, A. M., Robinson, V. L. & Goudreau, P. N. Two-component signal transduction. *Annu. Rev. Biochem.* **69**, 183–215 (2000).
- Ohta, N., Grebe, T. W. & Newton, A. in *Prokaryotic Development* (eds Brun, Y. V. & Shimkets, L. J.) 341–359 (ASM Press, Washington DC, 2000).
- Domian, I. J., Quon, K. C. & Shapiro, L. Cell type-specific phosphorylation and proteolysis of a transcriptional regulator controls the G1-to-S transition in a bacterial cell cycle. *Cell* **90**, 415–424 (1997).
- Quon, K. C., Marczyński, G. T. & Shapiro, L. Cell cycle control by an essential bacterial two-component signal transduction protein. *Cell* **84**, 83–93 (1996).
- Quon, K. C., Yang, B., Domian, I. J., Shapiro, L. & Marczyński, G. T. Negative control of bacterial DNA replication by a cell cycle regulatory protein that binds at the chromosome origin. *Proc. Natl Acad. Sci. USA* **95**, 120–125 (1998).
- Holtzendorff, J. et al. Oscillating global regulators control the genetic circuit driving a bacterial cell cycle. *Science* **304**, 983–987 (2004).
- Laub, M. T., Chen, S. L., Shapiro, L. & McAdams, H. H. Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle. *Proc. Natl Acad. Sci. USA* **99**, 4632–4637 (2002).
- Domian, I. J., Reisenauer, A. & Shapiro, L. Feedback control of a master bacterial cell-cycle regulator. *Proc. Natl Acad. Sci. USA* **96**, 6648–6653 (1999).
- Laub, M. T., McAdams, H. H., Feldblyum, T., Fraser, C. M. & Shapiro, L. Global analysis of the genetic network controlling a bacterial cell cycle. *Science* **290**, 2144–2148 (2000).
- Holtzendorff, J., Reinhardt, J. & Viollier, P. H. Cell cycle control by oscillating regulatory proteins in *Caulobacter crescentus*. *Bioessays* **28**, 355–361 (2006).
- Jenal, U. & Fuchs, T. An essential protease involved in bacterial cell-cycle control. *EMBO J.* **17**, 5658–5669 (1998).
- Jacobs, C., Domian, I. J., Maddock, J. R. & Shapiro, L. Cell cycle-dependent polar localization of an essential bacterial histidine kinase that controls DNA replication and cell division. *Cell* **97**, 111–120 (1999).
- Jacobs, C., Ausmees, N., Cordwell, S. J., Shapiro, L. & Laub, M. T. Functions of the CckA histidine kinase in *Caulobacter* cell cycle control. *Mol. Microbiol.* **47**, 1279–1290 (2003).
- Hecht, G. B., Lane, T., Ohta, N., Sommer, J. M. & Newton, A. An essential single domain response regulator required for normal cell division and differentiation in *Caulobacter crescentus*. *EMBO J.* **14**, 3915–3924 (1995).
- Wu, J., Ohta, N. & Newton, A. An essential, multicomponent signal transduction pathway required for cell cycle regulation in *Caulobacter*. *Proc. Natl Acad. Sci. USA* **95**, 1443–1448 (1998).
- Wheeler, R. T. & Shapiro, L. Differential localization of two histidine kinases controlling bacterial cell differentiation. *Mol. Cell* **4**, 683–694 (1999).
- Jacobs, C., Hung, D. & Shapiro, L. Dynamic localization of a cytoplasmic signal transduction response regulator controls morphogenesis during the *Caulobacter* cell cycle. *Proc. Natl Acad. Sci. USA* **98**, 4095–4100 (2001).
- Skerker, J. M., Prasol, M. S., Perchuk, B. S., Biondi, E. G. & Laub, M. T. Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. *PLoS Biol.* **3**, e334 (2005).
- Biondi, E. G. et al. A phosphorelay system controls stalk biogenesis during cell cycle progression in *Caulobacter crescentus*. *Mol. Microbiol.* **59**, 386–401 (2006).
- Meisenzahl, A. C., Shapiro, L. & Jenal, U. Isolation and characterization of a xylose-dependent promoter from *Caulobacter crescentus*. *J. Bacteriol.* **179**, 592–600 (1997).
- Iniesta, A. A., McGrath, P. T., Reisenauer, A., McAdams, H. H. & Shapiro, L. A phospho-signaling pathway controls the localization and activity of a protease complex critical for bacterial cell cycle progression. *Proc. Natl Acad. Sci. USA* **103**, 10935–10940 (2006).
- Hung, D. Y. & Shapiro, L. A signal transduction protein cues proteolytic events critical to *Caulobacter* cell cycle progression. *Proc. Natl Acad. Sci. USA* **99**, 13160–13165 (2002).
- Matroule, J. Y., Lam, H., Burnette, D. T. & Jacobs-Wagner, C. Cytokinesis monitoring during development; rapid pole-to-pole shuttling of a signaling protein by localized kinase and phosphatase in *Caulobacter*. *Cell* **118**, 579–590 (2004).
- Lam, H., Matroule, J. Y. & Jacobs-Wagner, C. The asymmetric spatial distribution of bacterial signal transduction proteins coordinates cell cycle events. *Dev. Cell* **5**, 149–159 (2003).
- Pierce, D. L. et al. Mutations in DivL and CckA rescue a *divJ* null mutant of *Caulobacter crescentus* by reducing the activity of CtrA. *J. Bacteriol.* **188**, 2473–2482 (2006).
- Wang, Y., Jones, B. D. & Brun, Y. V. A set of *ftsZ* mutants blocked at different stages of cell division in *Caulobacter*. *Mol. Microbiol.* **40**, 347–360 (2001).
- Wortinger, M., Sackett, M. J. & Brun, Y. V. CtrA mediates a DNA replication checkpoint that prevents cell division in *Caulobacter crescentus*. *EMBO J.* **19**, 4503–4512 (2000).
- Sciochetti, S. A., Lane, T., Ohta, N. & Newton, A. Protein sequences and cellular factors required for polar localization of a histidine kinase in *Caulobacter crescentus*. *J. Bacteriol.* **184**, 6037–6049 (2002).
- Pomeroy, J. R., Sontag, E. D. & Ferrell, J. E. Jr. Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nature Cell Biol.* **5**, 346–351 (2003).
- Collier, J., Murray, S. R. & Shapiro, L. DnaA couples DNA replication and the expression of two cell cycle master regulators. *EMBO J.* **25**, 346–356 (2006).
- Murray, A. W. & Kirschner, M. W. Dominoes and clocks: the union of two views of the cell cycle. *Science* **246**, 614–621 (1989).
- Cross, F. R. Two redundant oscillatory mechanisms in the yeast cell cycle. *Dev. Cell* **4**, 741–752 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y.S. Liu and K. Thorn for help with microscopy, and K. Thorn, A. Komeili, L. Garwin and A. Murray for discussions and/or comments on the manuscript. We acknowledge support from the Office of Science (BER), US Department of Energy (M.T.L.) and the US National Science Foundation (K.R.R.). Support was also provided in part by an NIH NIGMS Center of Excellence grant to Harvard University.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.T.L. (laub@mit.edu).

MARSIS radar sounder evidence of buried basins in the northern lowlands of Mars

Thomas R. Watters¹, Carl J. Leuschen², Jeffrey J. Plaut³, Giovanni Picardi⁴, Ali Safaeinili³, Stephen M. Clifford⁵, William M. Farrell⁶, Anton B. Ivanov³, Roger J. Phillips⁷ & Ellen R. Stofan⁸

A hemispheric dichotomy on Mars is marked by the sharp contrast between the sparsely cratered northern lowland plains and the heavily cratered southern highlands. Mechanisms proposed to remove ancient crust or form younger lowland crust include one or more giant impacts, subcrustal transport by mantle convection, the generation of thinner crust by plate tectonics, and mantle overturn following solidification of an early magma ocean^{1–7}. The age of the northern lowland crust is a significant constraint on these models. The Mars Advanced Radar for Subsurface and Ionospheric Sounding (MARSIS) instrument on the European Space Agency's Mars Express spacecraft is providing new constraints on the martian subsurface⁸. Here we show evidence of buried impact basins ranging in diameter from about 130 km to 470 km found over 14 per cent of the northern lowlands. The number of detected buried basins >200 km in diameter indicates that the lowland crust is ancient, dating back to the Early Noachian epoch. This crater density is a lower limit because of the likelihood that not all buried basins in the area surveyed by MARSIS have been detected. An Early Noachian age for the lowland crust has been previously suggested on the basis of a large number of quasi-circular topographic depressions interpreted to be evidence of buried basins^{9–11}. Only a few of these depressions in the area surveyed by MARSIS, however, correlate with the detected subsurface echoes. On the basis of the MARSIS data, we conclude that the northern lowland crust is at least as old as the oldest exposed highland crust. This suggests that the crustal dichotomy formed early in the geologic evolution of Mars.

In the immediate post-Viking era of Mars exploration, it was considered that the crust of the northern lowlands might be as young as the plains materials that cover it^{12–14}. The relatively young lowland plains and Late Noachian to Early Hesperian faulting along the dichotomy boundary led to models suggesting late-stage formation of the crustal dichotomy^{5,6}. Predictions from planetary accretion models and isotope data from the SNC (Shergottites, Nakhlites and Chassigny) meteorites suggest that the crust of Mars formed very early, within 50 Myr of the formation of the Solar System (see ref. 15). This presents a challenge even for models invoking early episodes of plate recycling¹⁶ and convection driven subcrustal transport^{17,18}, because these mechanisms require timescales of the order of hundreds of millions of years to form the dichotomy¹⁵.

The Mars Orbiter Laser Altimeter (MOLA) on Mars Global Surveyor revealed a population of craters¹⁹ and subdued quasi-circular depressions (QCDs)^{9–11} in the northern lowlands, interpreted to be ancient impact craters and basins buried by subsequent resurfacing. QCDs are identified as possible impact features on the basis of

concentric segments of topographic contours fitted to stretched MOLA data⁹. The areal density of QCDs suggests that the northern lowland crust is comparable in age to the exposed southern highlands, but younger than the highlands basement that includes highland QCDs¹¹. To date, there has been no independent confirmation (with the exception of the Utopia basin based on gravity and

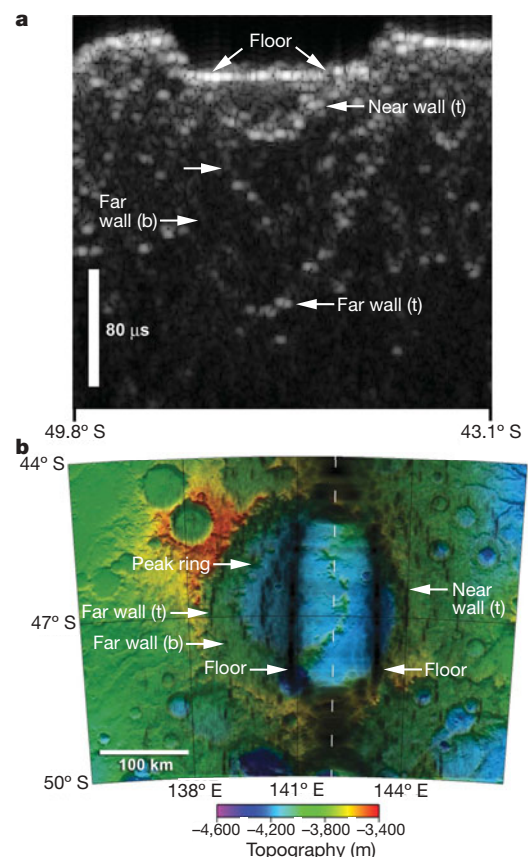


Figure 1 | MARSIS surface clutter simulation of the Kepler basin. **a**, **b**, The radargram (**a**) and the ground-range surface projection overlaid on MOLA topography (**b**) simulate the MARSIS 4 MHz band. The basin rim walls, both top (t) and bottom (b) from the near and far walls relative to the ground track, the basin floor, and the peak ring are sources of echoes (labelled). The white dashed line indicates the location of simulated nadir ground track.

¹Center for Earth and Planetary Studies, National Air and Space Museum, Smithsonian Institution, Washington DC 20560, USA. ²Center for Remote Sensing of Ice Sheets, The University of Kansas, Lawrence, Kansas 66045, USA. ³Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA. ⁴Infocom Department, "La Sapienza" University of Rome, 00184 Rome, Italy. ⁵Lunar and Planetary Institute, 3600 Bay Area Boulevard, Houston, Texas 77058, USA. ⁶NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ⁷Department of Earth and Planetary Sciences, Washington University, St Louis, Missouri 63130, USA. ⁸Proxemy Research, Laytonville, Maryland 20882, USA.

magnetic data²⁰) that QCDs in the lowlands are related to impact features. MARSIS is providing new data on the shallow crust of Mars⁸. We report here radar echoes obtained by MARSIS in the early phase of operations that indicate buried impact features in the northern lowlands. The MARSIS data collected thus far reveal 11 buried basins, many with no expression in imaging or topography.

The MARSIS instrument is a multi-frequency synthetic aperture orbital sounding radar that operates in four frequency bands between 1.3 MHz and 5.5 MHz in its subsurface sounding modes⁸ (Supplementary Note 1). The first months of MARSIS operations (June–July 2005) were over the northern lowlands. Much of the smooth plains of the northern lowlands consist of Late Hesperian and Early Amazonian age sedimentary surficial deposits of the Vastitas Borealis Formation^{21–23}. Partially buried by these surficial deposits are numerous wrinkle ridges similar to ridges found in exposed volcanic plains in the highlands¹⁹. Evidence of olivine-rich deposits associated with some fresh impact craters in the lowlands from the OMEGA (Observatoire pour la Minéralogie, l'Eau, les Glaces, et l'Activité) instrument on the Mars Express spacecraft suggests that the ridged plains beneath the Vastitas Borealis Formation sediments are volcanic in origin²⁴.

Time-delay renderings of MARSIS sounding data (radargrams) over the northern lowlands show parabolic-shaped echoes interpreted to be off-nadir clutter from buried features. Simulated radar reflections from the ~230-km-diameter Kepler impact basin using MOLA topography have the same distinctive parabolic shape (Fig. 1). Kepler exhibits many of the characteristics typical of impact basins of its size. The rim walls in some locations are broad and terraced, and the basin has a peak ring structure that is common in martian craters with diameters >45 km (ref. 25). Comparison of the parabolic echoes in the simulation radargram (Fig. 1a) with the echoes projected to ground-range geometry (Fig. 1b) indicates that the sources of the reflections are the top and bottom of the rim walls, the peak ring, and the basin floor. Because MARSIS receives surface and subsurface radar backscatter from the right and left of the ground track simultaneously, there is a left/right ambiguity, and thus echoes are projected on both sides of the track.

Adjacent orbits 1892 and 1903 over Chryse Planitia show multiple, complex parabolic echoes (Fig. 2a, b), analogous to the Kepler basin simulated echoes (Fig. 1). In ground-range projections of the radargrams, parabolic echoes project into arcs with constant curvature on the surface that can be used to infer the probable location and size of the basin (Fig. 2c, d). Overlapping echoes in the ground-range projections of the two radargrams indicate that returns are from the same subsurface feature, eliminating the left/right ambiguity. Circular fits to projected arcs suggest that an ~220-km-diameter basin is superimposed on a larger ~310-km-diameter basin (Supplementary Note 2). A prominent linear echo in orbit 1903 (Fig. 2a, c) is comparable to the simulated echo generated by the floor of Kepler basin (Fig. 1). The time delay corresponds to a depth of 2.0–2.5 km if the feature is directly below the spacecraft, and it may be the interface between the basin floor and the fill material⁸. Parabolic echoes in MARSIS orbits over Amazonis Planitia suggest buried basins beneath relatively young (Middle to Late Amazonian) lava flows. Ground-range projection of echoes shows that, as with orbits 1892 and 1903 (Fig. 2), arcs may consist of distinct segments with different radii of curvature. A parabolic echo in orbit 1886 (Fig. 3a) suggests an ~210-km-diameter basin that the orbit track crosses near the basin centre. A second discontinuous echo may be from a rim related structure, suggesting that the basin could be up to ~240 km in diameter. The ground-range projection of a prominent parabolic echo in orbit 1897 (Fig. 3b) is consistent with a buried basin ~140 km in diameter (Fig. 3d). Overall, buried basins have been detected in Chryse, Acidalia, Amazonis, Elysium and Utopia Planitiae (Table 1) (Supplementary Note 3).

The buried basins detected by MARSIS can be used to estimate an independent areal crater density ($N(200)$: a measure of the cumulative number of basins with diameters >200 km per 10^6 km^2) for the lowland crust. The area of the northern lowlands surveyed by MARSIS is estimated to be $\sim(4.63 \pm 1) \times 10^6 \text{ km}^2$ ($\sim 14 \pm 3\%$ of the lowlands) (Supplementary Note 4). This corresponds to an $N(200)$ crater density for the lowland crust of ~1.9 (range ~1.6–2.6), greater than the $N(200)$ crater density of Early Noachian boundary (1.28, based on $N(16)$ crater counts²⁶ extrapolated to larger

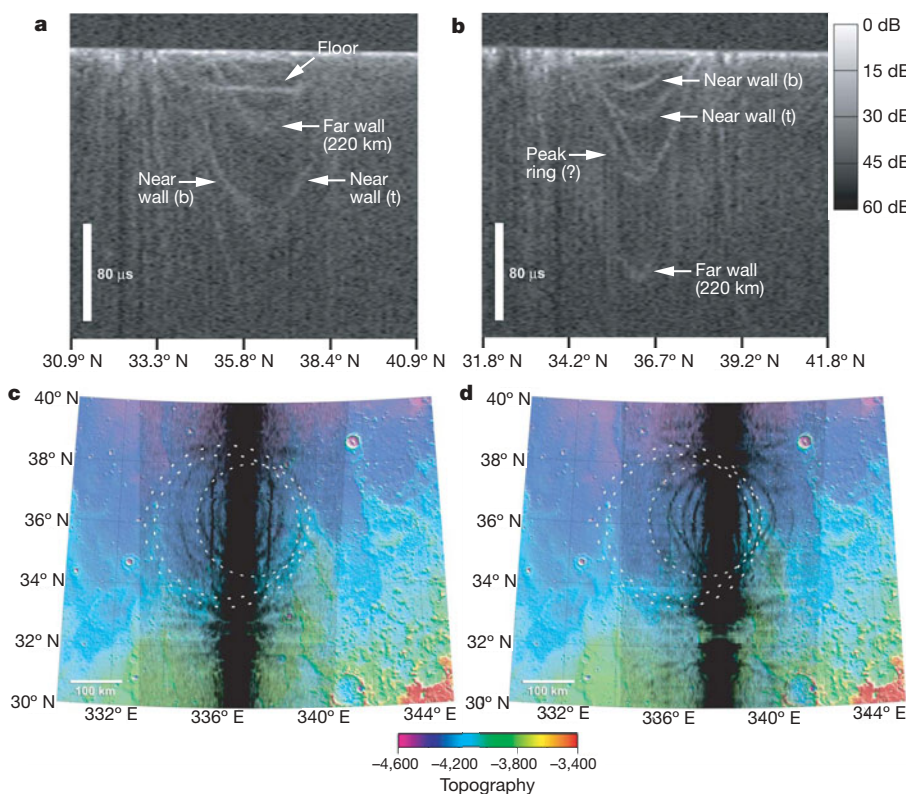


Figure 2 | Radargrams and ground-range projections of MARSIS data in Chryse Planitia.

a, b, Radargrams for MARSIS orbits 1903 (**a**) and 1892 (**b**) that cross Chryse Planitia. Echoes are plotted in time-delay versus position along the orbit track. The peak surface return is corrected to agree with the MOLA topography.

c, d, Ground-range projections of orbit 1903 (**c**) and 1892 (**d**) radargrams are overlaid on MOLA colour-coded shaded relief. Parabolic echoes that project as arcs on the surface are interpreted to be from the near and far rim walls (relative to the orbit track) of buried impact basins. The dashed white circles are approximate fits to the arcs (echoes labelled in radargrams). Echoes are interpreted to be from the far rim wall of an ~220-km-diameter basin (corresponding echoes in **a** and **b** labelled 'Far wall') superimposed on a larger ~310-km-diameter basin that may have echoes from the top and bottom of the near rim wall (corresponding echoes in **a** and **b** labelled 'Near wall').

diameters with a -2 power law). An Early Noachian age for the northern lowland crust is also suggested by a cumulative frequency plot of the MARSIS basins (Supplementary Fig. 1).

The MARSIS-based estimate of the crater density age of the lowlands crust is a lower limit because of the probability that not all the buried basins in the area surveyed have been detected. Several factors may contribute to the lack of subsurface scatter from buried impact basins, in particular the dielectric properties of the fill materials. The likelihood of detection is enhanced if basins are filled with materials that have a significant dielectric contrast with the crustal material. Conversely, basins filled with materials that are similar to the crustal material and homogeneously stratified may not be detected because the dielectric contrast is too small. Simulations indicate that the rim wall echo is sensitive to the rim wall slope. This suggests that there may be viewing geometry limitations and that only some combination of rim wall slope and spacecraft altitude results in a surface perpendicular to the instrument viewing angle. Other factors, like the degradation state and depth of burial, may also influence the detection of subsurface echoes from buried impact basins.

The QCDs in the northern lowlands with diameters >200 km give an $N(200)$ crater density of 2.5 (ref. 11), comparable to the MARSIS-based $N(200)$ crater density for the lowland crust. Many of the mapped QCDs¹¹ are within the area of the northern lowlands surveyed by MARSIS (Fig. 4). Some of these QCDs coincide with MARSIS basins. The location of a 315-km-diameter QCD ($\sim 35^\circ$ N, 336° E) roughly corresponds to the 310-km-diameter MARSIS basin in Chryse Planitia. A 470-km-diameter MARSIS basin in Acidalia Planitia approximately matches the location of an ~ 460 -km-diameter QCD ($\sim 49^\circ$ N, 341° E) (Fig. 4). Another QCD ~ 440 km in diameter ($\sim 56^\circ$ N, 14° E) corresponds approximately to the 470-km-diameter MARSIS basin in eastern Acidalia (Fig. 4). In Elysium Planitia, the location of the 200-km-diameter MARSIS basin

Table 1 | MARSIS buried basins in the northern lowlands

Location	Lat., Long. ($^\circ$ N, $^\circ$ E)	Diameter (km)	Orbits	Bands (MHz)
Chryse	36, 337	290–310	1903, 1892	3, 4
Chryse	36, 336	220	1903, 1892	3, 4
Acidalia	48, 342	400–470	1881	3
Acidalia	53, 339*	220	1881	3
Acidalia	50, 338	150	1903	3
Acidalia	54, 15*	470	1899	3
Amazonis	19, 208	210–240	1886	4
Amazonis	33, 206*	140	1897	4
Amazonis	13, 208*	210	1875	4
Elysium	6, 144*	120–200	1872	4
Utopia	16, 112*	270	1887	3

The MARSIS subsurface sounding bands are centred at 1.8, 3.0, 4.0 and 5.0 MHz.

* Basin centre location is not uniquely determined.

approximately coincides with a 220-km-diameter QCD ($\sim 6^\circ$ N, 145° E) (Fig. 4). After minimizing the offset by allowing for the left/right ambiguity in the basin location, the offset between the centres of the MARSIS basins and the QCDs ranges from ~ 60 km to >120 km. Most of the buried basins detected by MARSIS have no discernable topographic expression in MOLA-based shaded relief maps (Fig. 3, Supplementary Note 3). The agreement between QCDs and the MARSIS basins detected thus far supports the interpretation that QCDs are the surface expression of buried basins. However, the small number of QCDs that correlate with detected subsurface echoes may indicate that not all QCDs are related to buried basins.

The age of the northern lowland crust is critical in determining when the hemispheric dichotomy formed, and in constraining models for its origin. The results of this study indicate that the northern lowland crust is ancient, in agreement with the QCD-based crater density age^{9–11}, and suggest that the dichotomy formed in the Early

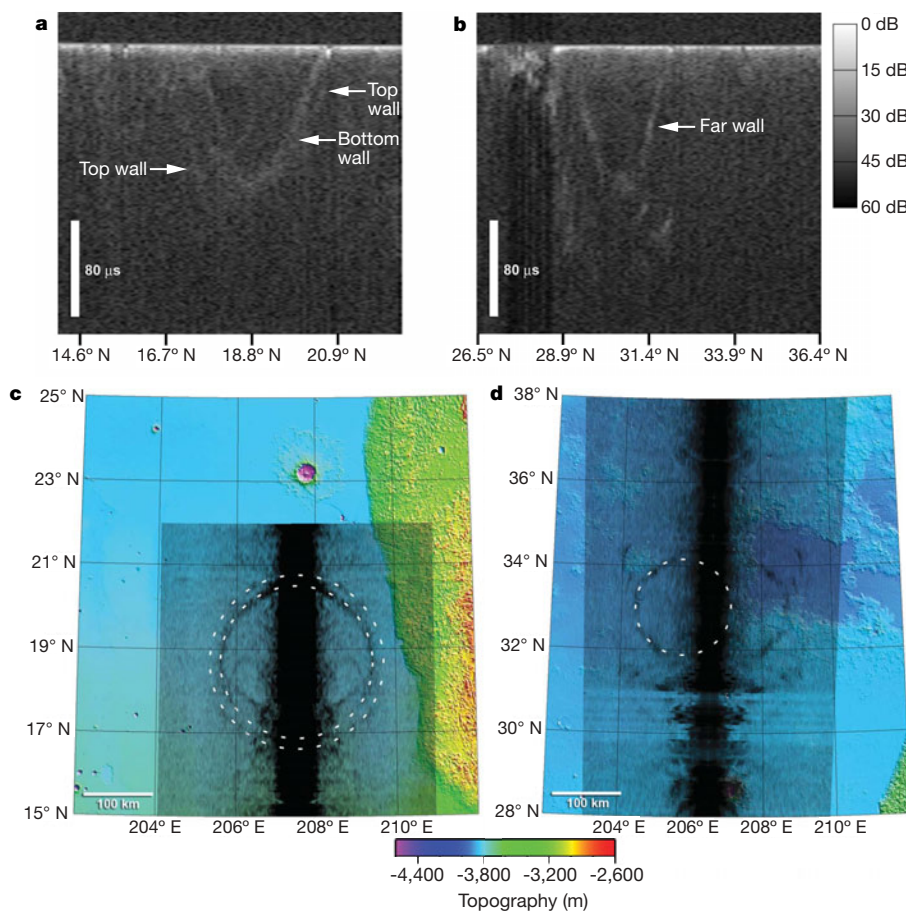


Figure 3 | Radargrams and ground-range projections of MARSIS data in Amazonis Planitia. **a, b**, Radargrams showing MARSIS data for orbits 1886 (**a**) and 1897 (**b**) that cross Amazonis Planitia. **c, d**, Ground-range projections of radargrams for 1886 (**c**) and 1897 (**d**) are overlaid on MOLA colour-coded shaded relief. The dashed white circles are approximate fits to the arcs (echoes labelled in radargrams). Echoes in 1886 (**c**) are interpreted to be from the top and bottom rim wall of an ~ 210 -km-diameter basin (corresponding echoes labelled in **a**). The echo in 1897 (**d**) is interpreted to be from the far rim wall of an ~ 140 -km-diameter basin (corresponding echo labelled in **b**). The best fit to the echo indicates that the orbit track is offset from the basin centre, thus the left/right ambiguity in the data does not allow a unique determination of the basin centre.

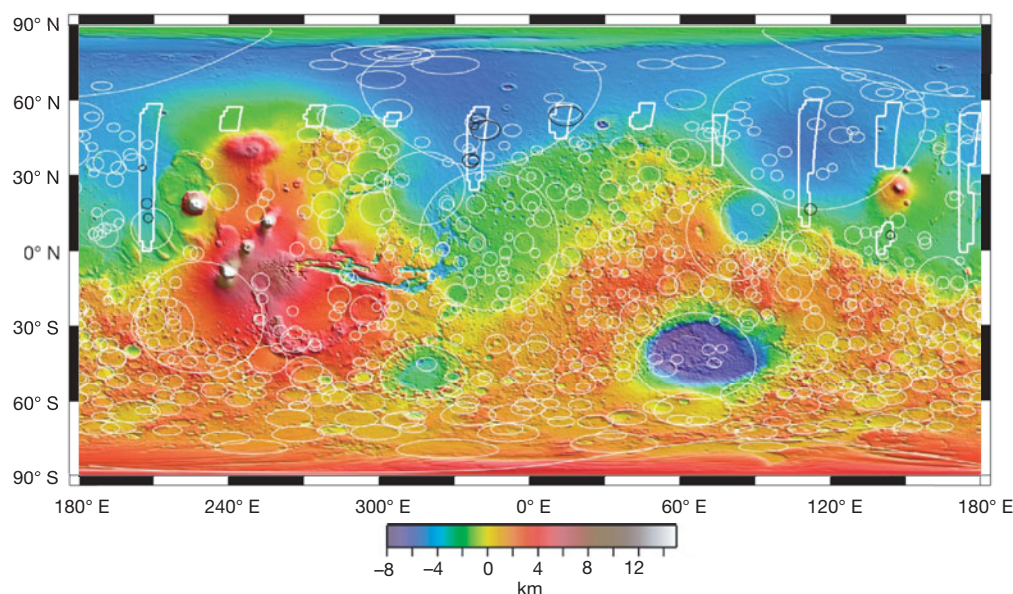


Figure 4 | Locations of MARSIS buried basins. The locations and inferred diameters from MARSIS echoes are shown in black on MOLA colour-coded shaded relief. The locations and diameters of QCDs with diameters >200 km mapped in ref. 11 are plotted in white. White polygons show the area covered by MARSIS orbits (see text). Where there is left/right ambiguity in the location of a MARSIS basin that coincides with a QCD, the plotted location reflects the minimum offset between the centres of the MARSIS basin and the QCD.

Noachian. The question of how early the hemispheric dichotomy formed may be addressed if the difference in age between the highland and lowland crust can be determined. The QCD-based $N(200)$ crater density for the highlands is estimated to be ~ 4.5 (exposed basins and QCDs) with an upper limit of ~ 8.5 based on extrapolation of the largest highland basins (exposed basins and QCDs)¹¹. The larger QCD-based $N(200)$ crater density suggests that the highland crust is older than the lowland basement. This areal crater density, however, includes many suspected impact basins that can not be independently verified by MARSIS sounding data or other geophysical data. Another approach to evaluating the age difference between the highland and lowland crust is by comparing the MARSIS-based crater density for the lowland basement to the crater density of the exposed highland basement. One of the oldest and best preserved areas of highland basement, in the area of Claritas Fossae ($\sim 28^\circ$ S, 259° E), has an $N(16)$ crater density of 294 ± 81 (ref. 26). This corresponds to an $N(200)$ density of $\sim 1.9 \pm 0.5$ (based on extrapolation to larger diameters using a -2 power law), comparable to the MARSIS-based $N(200)$ crater density of the lowland basement. Thus, we conclude from the MARSIS data that the lowland crust is at least as old as the oldest exposed highland crust. Although the MARSIS-based $N(200)$ crater density for the lowlands and the extrapolated $N(200)$ density for the highlands are lower limits, the similarity in age suggests that the crustal dichotomy formed early in the geologic evolution of Mars.

Received 9 June; accepted 12 October 2006.

- Wilhelms, D. E. & Squyres, S. W. The martian hemispheric dichotomy may be due to a giant impact. *Nature* **309**, 138–140 (1984).
- McGill, G. E. Buried topography of Utopia, Mars: Persistence of a giant impact depression. *J. Geophys. Res.* **94**, 2753–2759 (1989).
- Frey, H. & Schultz, R. A. Large impact basins and the mega-impact origin for the crustal dichotomy on Mars. *Geophys. Res. Lett.* **15**, 229–232 (1988).
- Wise, D. U., Golombek, M. P. & McGill, G. E. Tharsis province of Mars: Geologic sequence, geometry and a deformation mechanism. *Icarus* **38**, 456–472 (1979).
- McGill, G. E. & Dimitriou, A. M. Origin of the martian global dichotomy by crustal thinning in the late Noachian or early Hesperian. *J. Geophys. Res.* **95**, 12595–12605 (1990).
- Sleep, N. H. Martian plate tectonics. *J. Geophys. Res.* **99**, 5639–5656 (1994).
- Elkins-Tanton, L. T., Parmentier, E. M. & Hess, P. C. The formation of ancient crust on Mars through magma ocean processes. *J. Geophys. Res.* **110**, E12501, doi:10.1029/2005JE002480 (2005).
- Picardi, G. *et al.* Radar soundings of the subsurface of Mars. *Science* **310**, 1925–1928 (2005).
- Frey, H. V., Roark, J. H., Shockey, K. M., Frey, E. L. & Sakimoto, S. E. H. Ancient lowlands on Mars. *Geophys. Res. Lett.* **29**, doi:10.1029/2001GL013832 (2002).
- Frey, H. V. Impact constraints on the age of origin of the lowlands of Mars. *Geophys. Res. Lett.* **33**, doi:10.1029/2005GL024484 (2006).
- Frey, H. V. Impact constraints on, and a chronology for, major events in early Mars history. *J. Geophys. Res.* **111**, doi:10.1029/2005JE002449 (2006).
- Scott, D. H. & Tanaka, K. L. Geologic map of the western equatorial region of Mars, scale 1:15,000,000. *US Geol. Surv. Misc. Invest. Ser. Map I-1802-A*, (1986).
- Greeley, R. & Guest, J. E. Geologic map of the eastern equatorial region of Mars, scale 1:15,000,000. *US Geol. Surv. Misc. Invest. Ser. Map I-1802-B*, (1987).
- Tanaka, K. L. & Scott, D. H. Geologic maps of the polar regions of Mars, scale 1:15,000,000. *US Geol. Surv. Misc. Invest. Ser. Map I-1802-C*, (1987).
- Solomon, S. C. *et al.* New perspectives on ancient Mars. *Science* **307**, 1214–1220 (2005).
- Lenardic, A., Nimmo, F. & Moresi, L. Growth of the hemispheric dichotomy and the cessation of plate tectonics on Mars. *J. Geophys. Res.* **109**, E02003, doi:10.1029/2003JE002172 (2004).
- Zhong, S. & Zuber, M. T. Degree-1 mantle convection and the crustal dichotomy on Mars. *Earth Planet. Sci. Lett.* **189**, 75–84 (2001).
- Roberts, J. H. & Zhong, S. Degree-1 convection in the Martian mantle and the origin of the hemispheric dichotomy. *J. Geophys. Res.* **111**, doi:10.1029/2005JE002668 (2006).
- Head, J. W., Kreslavsky, M. A. & Pratt, S. Northern lowlands of Mars: Evidence for widespread volcanic flooding and tectonic deformation in the Hesperian Period. *J. Geophys. Res.* **107**, doi:10.1029/2000JE001445 (2002).
- Zuber, M. T. The crust and mantle of Mars. *Nature* **412**, 220–227 (2001).
- Tanaka, K. L. *et al.* Resurfacing history of the northern plains of Mars based on geologic mapping of Mars Global Surveyor data. *J. Geophys. Res.* **108** (E4), doi:10.1029/2002JE001908 (2003).
- Tanaka, K. L. *et al.* Geologic map of the northern plains of Mars. *US Geol. Surv. Misc. Invest. Ser. Map I-2888*, (2005).
- Kreslavsky, M. A. & Head, J. W. The fate of outflow channel effluents in the northern lowlands of Mars: The Vastitas Borealis Formation as a sublimation residue from frozen ponded bodies of water. *J. Geophys. Res.* **107** (E12), 5121, doi:10.1029/2001JE001831 (2002).
- Mustard, J. F. *et al.* Olivine and pyroxene diversity in the crust of Mars. *Science* **307**, 1594–1597 (2005).
- Melosh, H. J. *Impact Cratering: A Geologic Process* (Oxford Univ. Press, New York, 1989).
- Tanaka, K. L. The stratigraphy of Mars. *J. Geophys. Res.* **91**, E139–E158 (1986).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. W. Head and D. D. Blankenship for constructive criticism and advice that greatly improved the paper. We also thank H. V. Frey for providing the locations and diameters of the QCDs. We are grateful to the other members of the MARSIS Science team and to the instrument operation staff. MARSIS is managed by the Agenzia Spaziale Italiana (ASI) and the National Aeronautics and Space Administration (NASA). The Mars Express mission is managed and operated by the European Space Agency. The research activities of the MARSIS PI and Italian investigators were supported by the Mars Express/ASI programme; the work of the US investigators was supported by the Mars Express/NASA project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.R.W. (watterst@si.edu).

Minority spin condensate in the spin-polarized superfluid ^3He A_1 phase

A. Yamaguchi¹, S. Kobayashi¹, H. Ishimoto¹ & H. Kojima²

The magnetic properties of ^3He in its various phases originate from the interactions among the nuclear spins¹. The spin-polarized ‘ferromagnetic’ superfluid ^3He A_1 phase² (which forms below 3 mK between two transition temperatures, T_{c1} and T_{c2} , in an external magnetic field) serves as a material in which theories of fundamental magnetic processes and macroscopic quantum spin phenomena may be tested. Conventionally, the superfluid component of the A_1 phase is understood^{3–6} to contain only the majority spin condensate, having energetically favoured paired spins directed along the external field and no minority spin condensate having paired spins in the opposite direction. Because of difficulties in satisfying both the ultralow temperature and high magnetic field required to produce a substantial phase space, there exist few studies of spin dynamics phenomena that could be used to test the conventional view of the A_1 phase. Here we develop a mechanical spin density detector that operates in the required regime, enabling us to perform measurements of spin relaxation in the A_1 phase as a function of temperature, pressure and magnetic field. Our mechanical spin detector is based in principle on the magnetic fountain effect⁷; spin-polarized superfluid motion can be induced both magnetically and mechanically, and we demonstrate the feasibility of increasing spin polarization by a mechanical spin filtering process. In the high temperature range of the A_1 phase near T_{c1} , the measured spin relaxation time is long, as expected^{2,8,9}. Unexpectedly, the spin relaxation rate increases rapidly as the temperature is decreased towards T_{c2} . Our measurements, together with Leggett–Takagi theory⁵, demonstrate that a minute presence of minority spin pairs is responsible for this unexpected spin relaxation behaviour. Thus, the long-held conventional view² that the A_1 phase contains only the majority spin condensate is inadequate.

The dipolar interaction between nuclear spins in bulk normal liquid ^3He results in an intrinsic longitudinal spin relaxation time⁹ that diverges at low temperature (T) as T^{-2} . Magnetic relaxation of liquid ^3He in any experiment at low temperatures, however, inevitably involves both the intrinsic relaxation and an additional—usually much more rapid—extrinsic relaxation that occurs at the surfaces of the apparatus or particles introduced into it¹⁰. The magnetic fluctuations in the dense solid-like ^3He layer adjacent to surfaces are thought to be responsible for the magnetic relaxation time linearly proportional to T and to magnetic field H (refs 11, 12). Below the superfluid transition temperatures, new mechanisms of intrinsic magnetic relaxation mediated by the condensate spin pairs become possible^{4,5}. Measurements¹³ of NMR linewidths in the ^3He A phase in small magnetic fields have been interpreted in terms of such intrinsic relaxation mechanisms. We present the first (to our knowledge) measurements and analysis of the intrinsic spin relaxation using an unusual mechanical spin density detector in the bulk ^3He A_1 phase in high magnetic fields.

Consider a small detector chamber connected to a much larger reservoir chamber filled with ^3He A_1 phase via a ‘superleak’ (a very narrow constriction; see Fig. 1 and Supplementary Information). A rapid increase (or decrease) in spin polarization in the small chamber is generated by superfluid flow (‘superflow’) induced by an applied magnetic field gradient across the superleak. The superflow in the superleak (spin filter) carries the spin-polarized superfluid component of A_1 phase. If the superleak is ideal, the spin polarization gradient across it, in the absence of spin relaxation, would be balanced by a steady magnetic fountain pressure difference between the two chambers⁷. In the presence of spin relaxation, however, the pressure difference relaxes, and the observed fountain pressure relaxation gives a direct measure of the spin polarization relaxation.

The principal mechanism that governs the quasistatic (superfluid acceleration is ignored) response of our detector is that the superfluid maintains equal chemical potential at both ends of the superleak. This leads to the magnetic fountain effect⁷ expressed by: $\delta P/\rho + (\gamma\hbar/2m)(\gamma\delta S/\chi - \delta H) = 0$, where δP , δS and δH are the differential pressure, spin density and magnetic field across the superleak, respectively, m is the mass of ^3He atom, ρ the mass density, γ the

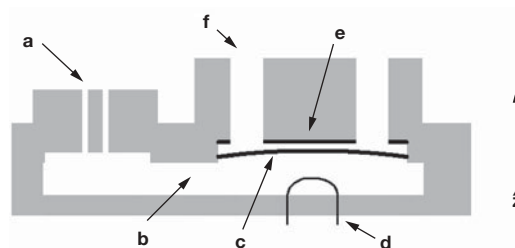


Figure 1 | Schematic cross-sectional view of our spin density detector. **a**, Superleak channels (two parallel channels made by etching out 18 μm thick, 3.0 mm long, and 3.6 mm wide aluminium foils cast into Stycast 1266 epoxy). **b**, Detector chamber of volume 0.3 cm^3 . **c**, Lightly stretched circular flexible membrane with a tension $1.9 \times 10^5 \text{ dyn cm}^{-1}$ (9 mm diameter and 9 μm thick aluminium-coated Mylar membrane). **d**, Vibrating wire viscometer. **e**, Stationary metal film electrode. **f**, Four vent holes (1.5 mm diameter and 4 mm long, only two are shown). The working principle is based on the magnetic fountain effect in ^3He A_1 phase. The whole detector assembly is immersed in a large reservoir of liquid ^3He in a uniform static magnetic field. The static and gradient magnetic fields are applied along the $-\hat{z}$ (down) direction parallel to the superleak channels. The displacement of the flexible diaphragm is detected by measuring the changes in the capacitance between the fixed electrode and the diaphragm. The displacement gives a measure of the induced (differential) magnetic fountain pressure across the superleak. The diaphragm is located on the same side of the detector chamber, and the vent holes are provided such that the measured differential pressure reflects more accurately that across the superleak. All structural parts except those noted are fabricated with Stycast 1266 epoxy. Electrical leads are not shown.

¹Institute for Solid State Physics, University of Tokyo, Kashiwa, Chiba 277-8581, Japan. ²Serin Physics Laboratory, Rutgers University, Piscataway, New Jersey 08854, USA.

gyromagnetic ratio, and χ the magnetic susceptibility¹. The induced differential pressure δP is related to the displacement Z of the flexible membrane (of area A_m and tension σ) by $8\pi\sigma Z = A_m\delta P$. The membrane displacement is a direct measure of the magnetically induced pressure. Since the superleak is imperfect, the induced pressure produces small concurrent normal fluid flows. The measured membrane dynamics is determined by both the relaxation (with relaxation time T_1) in spin density and the normal fluid flow (with relaxation time τ_n). These considerations, when taken together with the conservation of mass and the spin relaxation, lead to the membrane relaxation time τ :

$$\frac{1}{\tau} = \left(\frac{1}{\tau_n} + \frac{\alpha}{T_1} \right) \left(\frac{1}{1+\alpha} \right) \quad (1)$$

where τ_n is the normal fluid flow relaxation time, $\alpha = 8\pi\sigma\chi V / (\hbar\gamma\rho A_m/2m)^2$ ($=24$) is the mechanical to magnetic energy density ratio and V is the volume of the detector chamber (see Supplementary Information).

Examples of the membrane displacement are shown in Fig. 2 when a magnetic field difference $\delta H \approx 0.63$ mT is applied across the superleak within a typical time of 100 ms. As expected, there is no measurable response in the normal ($T > 2.50$ mK) and A_2 ($T < 2.11$ mK) phases. Within the A_1 phase, the displacement increases linearly in time to a maximum, which depends on the applied field difference and the field ramp rate. For a large enough field difference, dZ/dt tends to saturate, presumably limited by a critical velocity (measured to be ~ 0.5 mm s⁻¹, consistent with others¹⁴) in the superleak. After reaching the maximum, the displacement decays exponentially in time at all temperatures except near T_{c1} . The measured relaxation time does not depend on the maximum displacement amplitude.

The decaying portion of the measured response is fitted with an exponential function with time constant τ . Figure 3 shows τ as a function of normalized reduced temperature, $r = (T_{c1} - T) / (T_{c1} - T_{c2})$, where $T_{c1} - T_{c2} = 0.052$ H mK T⁻¹, at a pressure of 21 bar in static magnetic fields up to 8 T. Non-exponential and erratic

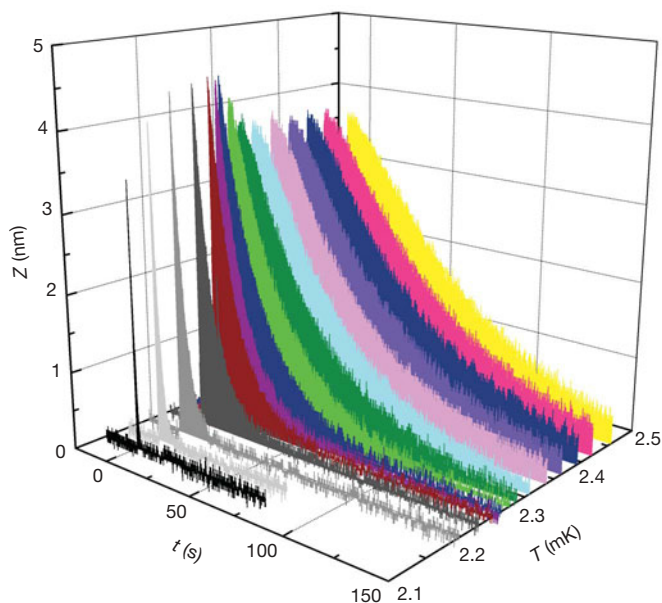


Figure 2 | Time response of the detector membrane displacement Z to an applied step in field gradient at time $t = 0$. The liquid pressure is 21 bar and the applied static field is 8 T. The colours distinguish 14 temperatures (see Supplementary Information) at which data are shown. Many more data were taken than shown. There is no response outside the A_1 phase temperature range. The onset temperatures of the magnetic fountain effect coincide with the kinks in shear viscosity²² at T_{c1} and T_{c2} as detected by the viscometer placed in the detector chamber. After the initial rise, the response decays exponentially.

relaxations near T_{c1} , where $r < 0.18$, are excluded from Fig. 3. Clearly, the observed relaxation time is distinct from that in normal fluid. At a given field, τ gradually decreases as the temperature decreases from $r = 0$ and it rapidly decreases over a relatively small temperature interval at a 'kink' temperature (indicated by an arrow). The kink temperature tends towards T_{c1} as the static field is decreased towards zero. The relaxation time finally tends towards zero as the temperature approaches T_{c2} ($r = 1$).

At a given temperature r , τ increases as the static field is increased. The presence of the 'kinks' introduces more complication to the field dependence at low fields and small values of r . However, it is clear that τ tends to increase with increasing applied static field at a given value of r . Except in the region of kinks, τ increases linearly with applied field above 1 T (see Supplementary Information). This field dependence is reminiscent of the spin relaxation mediated by the solid-like ³He layer at the boundaries observed in magnetic relaxation in liquid ³He immersed in fluorocarbon particles¹¹, but this is not the case, as discussed below.

Instead of using a magnetic field gradient, spin polarization may be induced by mechanical pumping by applying a step voltage to the flexible membrane with respect to the fixed electrode, thereby drawing in spin-polarized superfluid into the chamber. The process is in effect spin pumping through the superleak spin filter. Studies of time response to such mechanical manipulation of the spin-polarized superfluid show that the observed relaxation time is the same as those observed subsequent to magnetically induced spin flow under the same conditions as in Fig. 3. This observation proves that the mechanically driven superflow induces spin polarization and that the relaxation time is not dependent on the presence of applied magnetic field gradient.

An important check to distinguish the source of the observed spin relaxation as the bulk liquid or the solid-like boundary layer at the walls is to replace the ³He boundary layer by magnetically inert ⁴He, which preferentially covers the boundary walls¹⁰. In a separate experiment, a sufficient amount of ⁴He was introduced to cover all surfaces with five monolayers of ⁴He, which should remove the magnetically active layer (see ref. 15, for example). The observed relaxation times with the ⁴He coverage are the same within 15% as those in pure ³He at all temperatures and magnetic fields. This provides strong evidence that the observed spin relaxation occurs within the bulk liquid.

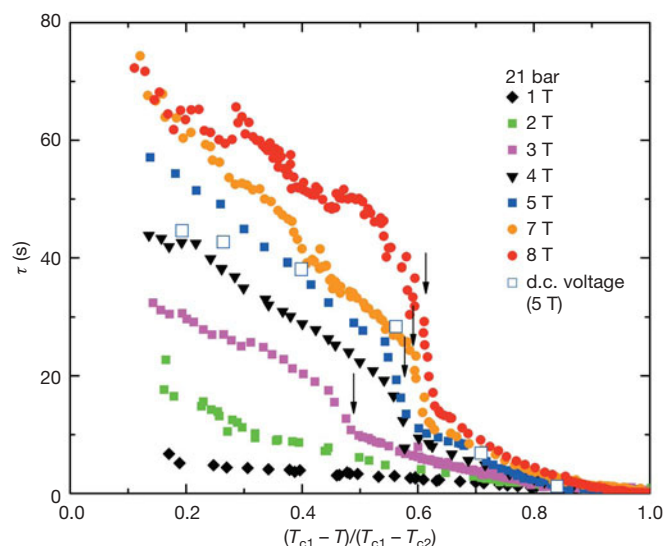


Figure 3 | Measured relaxation time versus normalized reduced temperature at 21 bar. Relaxation time τ tends to vanish as T_{c2} is approached, increases as T_{c1} is approached, and shows a sharp change at an intermediate temperature that depends on the applied static field. Open squares show the measured relaxation after spin polarization was induced by a d.c. voltage applied to the membrane (see text) at a static field of 5 T.

Surprisingly, the measured relaxation time tends to vanish as T_{c2} is approached at all magnetic fields. To highlight the region near T_{c2} , and to make direct comparison with theory, T_1^{-1} is extracted from the measured the relaxation rate τ^{-1} using equation (1). The result is plotted against $t = (T - T_{c2})/T_{c2}$ in Fig. 4 for the data at 2 T. To our knowledge, no spin transport mechanism has been predicted to show such an increase in spin relaxation within the A_1 phase near T_{c2} . Below, we interpret the behaviour near T_{c2} with the theory by Leggett and Takagi⁵ and the predicted (by Monien and Tewordt^{16,17}) presence of a tiny amount of minority spin condensate (MSC) in the A_1 phase.

Leggett and Takagi⁵ introduced a two-fluid model of spin relaxation in superfluid ^3He . Very briefly, the total spin is decomposed into those spins carried by the normal and superfluid components. When the spin polarization of the superfluid component changes, its associated magnetic field (or chemical potential) can change essentially instantaneously (within a timescale of $\hbar/\Delta$, where Δ is the energy gap^{2,3}) with its magnetic susceptibility at constant normal component spin. This induces a chemical potential difference between the superfluid and normal fluid components. The spins in the two components approach equilibrium by spin-conserving collisions with a characteristic relaxation time. During these two processes, the spin polarization and field can get out of phase and produce dissipation and an associated spin relaxation rate.

According to this Leggett and Takagi theory, the relaxation rate of longitudinal magnetization in the A phase is given by:

$$1/T_1 = Y_2 \tau_n \chi \Omega_{\parallel}^2 / [(1 - Y_2) \chi_0] \quad (2)$$

where Y_2 is the Yoshida function of the second kind¹⁸, $\chi/\chi_0 = (1 + Z_0/4)^{-1}$, χ_0 the magnetic susceptibility in the absence of Fermi liquid effects¹, Z_0 a Landau parameter¹, τ_n the quasiparticle relaxation time at T_c , and $\Omega_{\parallel}$ the longitudinal magnetic resonance frequency. Comparison of equation (2) with experiment has been difficult owing to the suspected presence of spin currents. Qualitative agreement, however, has been seen in the measurements of longitudinal resonance line width¹³ and of the longitudinal relaxation time¹⁹. Leggett and Takagi conjectured that the mechanism described above would be applicable to the A_1 phase if the changes in magnetic fields and polarizations were small compared to their original values.

An applied magnetic field H with associated field energy $\eta'H$, where η' is a constant¹⁶, tends to create A_1 phase with $\uparrow\uparrow$ spin pairs with an associated energy gap $\Delta_{\uparrow\uparrow}$ in accordance with particle-hole

asymmetry. This is the 'conventional' description of A_1 phase and there is no longitudinal magnetic resonance, that is, $\Omega_{\parallel} = 0$. Equation (2) then implies that T_1 would diverge, in disagreement with experiment. However, Monien and Tewordt¹⁶ show that inclusion of dipolar interaction with energy scale g_D tends to create MSC with opposite $\downarrow\downarrow$ spin pairs with energy gap $\Delta_{\downarrow\downarrow}$. Monien and Tewordt find that $\Delta_{\downarrow\downarrow}/\Delta_{\uparrow\uparrow} \approx g_D/\eta'H \approx 10^{-4}/H$, where H is in units of T. The presence of this minute MSC population implies a finite longitudinal NMR frequency, contrary to the 'conventional' description of the A_1 phase. We apply this new $\Omega_{\parallel}$ to equation (2). The minuscule MSC population turns out to produce a huge influence on spin relaxation.

To apply the Leggett and Takagi theory to the A_1 phase, the function Y_2 is transformed to $Y_2^{A_1}$ as follows: in the tabulation¹⁸ of Y_2 , we set $Y_2(T/T_c=1) = Y_2^{A_1}(r=1)$ and $Y_2(T/T_c=(T_{c1}-T_{c2})/T_{c1}) = Y_2^{A_1}(r=0)$ for given H , and interpolating in between by setting $Y_2(T/T_c) = Y_2^{A_1}(r)$. The parallel resonance frequency $\Omega_{\parallel}$ in the A_1 phase in 2 T is taken from Fig. 9 of Monien and Tewordt¹⁶. The evaluated relaxation rate from equation (2) with no adjusted parameter is shown in Fig. 4. The measured and theoretical relaxation rates, where $t < 0.01$, are within a factor of four. The temperature dependence of the theoretical T_1^{-1} is quite similar to that of the data. The theoretical relaxation rate critically depends on the calculated longitudinal resonance frequency in the A_1 phase. In view of the unknown influence on the parallel resonance frequency by texture and detector chamber geometry, the agreement between theory and experiment may be considered good. Comparison between the measured dependences of T_1^{-1} on magnetic field and pressure to those expected from the theory show equally good agreement (see Supplementary Information).

Osheroff and Anderson²⁰ measured NMR in both the A_1 and A_2 phases in the vicinity of T_{c2} in 1.5 T. They concluded that no longitudinal resonance occurred in A_1 . The predicted¹⁶ longitudinal frequency of 2.5 kHz in the A_1 phase was unlikely to be detectable in their experiment. On the other hand, an increasing longitudinal relaxation rate $1/T_1$ was seen, in agreement with our observations, in the A_1 phase as T_{c2} was approached in a field of 0.305 T (ref. 19). It has been observed that the spin-entropy (second sound) wave propagation²¹ suffers an extra 'anomalous attenuation' in the vicinity of T_{c2} . The presence of MSC is likely to be the origin of this extra attenuation.

Our conclusion that the A_1 phase contains an MSC implies that it is an incipient A_2 phase, which would not support a magnetic fountain effect or a spin-entropy wave. Why then does the presence of MSC increase the spin relaxation rate but not suppress the magnetic fountain effect? To address this question, we estimate the pair breaking critical velocity² for the MSC, $v_c \approx \Delta_{\downarrow\downarrow}/p_F$, where p_F is the Fermi momentum¹ and $\Delta_{\downarrow\downarrow}$ is estimated according to Monien and Tewordt (see above). At 21 bar and 5 T, we find v_c is at most of the order of $1 \mu\text{m s}^{-1}$. The superflow velocity (greater than $50 \mu\text{m s}^{-1}$) in our superleak channels during ramping up of the applied magnetic field gradient certainly exceeds v_c . When the MSC is broken by the flow, the conventional A_1 phase is restored along with the magnetic fountain effect. The pair breaking velocity increases as the magnetic field is decreased, and this might be contributing in part to the increasing deviation between the predictions of the Leggett and Takagi theory and the relaxation time data at low magnetic fields (see Supplementary Fig. 3). The small MSC pair breaking velocity was very probably exceeded during the oscillatory period of superflow (typically $10 \mu\text{m s}^{-1}$) in spin-entropy wave experiments²¹. All previous spin-entropy wave observations have been made at frequencies greater than 10^2 Hz , which exceeds the spin relaxation rate of at most 20 s^{-1} observed here. Thus, once the system becomes supercritical, it tends to remain essentially in the conventional A_1 phase state, allowing spin-entropy wave propagation in the 'high' frequency measurements. Magnetic fountain effect experiments have uncovered novel phenomena previously undetected in the spin-polarized superfluid ^3He .

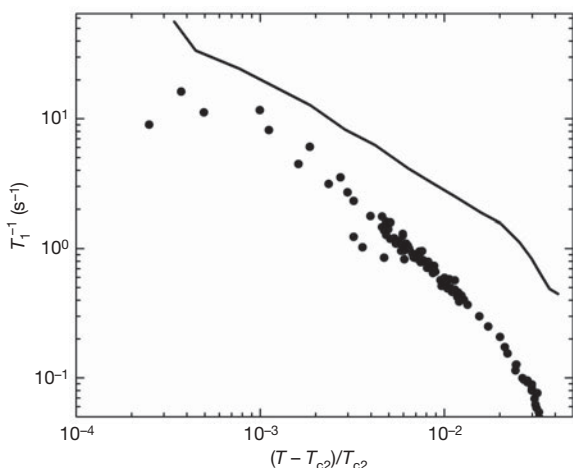


Figure 4 | Dependence of longitudinal relaxation rate on reduced temperature relative to T_{c2} . The dots show data evaluated from the measured relaxation time using equation (1); the line shows calculated T_1^{-1} , based on the presence of minority spin condensate¹⁶ and the Leggett and Takagi theory⁵ of spin relaxation, without any parameter adjustment (see text).

Received 9 June; accepted 25 October 2006.

1. Dobbs, E. R. *Helium Three* (Oxford Univ. Press, Oxford, 2000).
2. Vollhardt, D. & Wölfle, P. *The Superfluid Phases of Helium-3* (Taylor and Francis, London, 1990).
3. Wilks, J. & Betts, D. S. *An Introduction to Liquid Helium* (Oxford Univ. Press, Oxford, 1987).
4. Wheatley, J. C. Experimental properties of superfluid ^3He . *Rev. Mod. Phys.* **47**, 415–470 (1975).
5. Leggett, A. J. & Takagi, S. Orientational dynamics of superfluid ^3He : A “two-fluid” model. I. Spin dynamics with relaxation. *Ann. Phys.* **106**, 79–133 (1977).
6. Leggett, A. J. A theoretical description of the new phases of liquid ^3He . *Rev. Mod. Phys.* **47**, 331–414 (1975).
7. Liu, M. Broken relative symmetry and dynamics of the A_1 phase of ^3He . *Phys. Rev. Lett.* **43**, 1740–1743 (1979).
8. Devegvar, P. G. N. Spin dynamics of superfluid ^3He - A_1 . *Phys. Rev. B* **30**, 6349–6359 (1984).
9. Vollhardt, D. & Wölfle, P. Theory of spin relaxation by magnetic dipole interaction in a Fermi liquid: Application to ^3He . *Phys. Rev. Lett.* **47**, 190–193 (1981).
10. van Steenberg, A. S., Wiegers, S. A. J., Perenboom, J. A. A. J. & Maan, J. C. New surface relaxation mechanism for liquid ^3He in ^4He . *Phys. Rev. Lett.* **79**, 115–118 (1997).
11. Hammel, P. C. & Richardson, R. C. Relaxation of nuclear magnetization of liquid ^3He in confined geometries. *Phys. Rev. Lett.* **52**, 1441–1444 (1984).
12. Schuhl, A., Maegawa, S., Meisel, M. W. & Chapellier, M. Static and dynamic magnetic properties of ^3He confined by fluorocarbon microspheres. *Phys. Rev. B* **36**, 6811–6836 (1987).
13. Gully, W. J., Gould, C. M., Richardson, R. C. & Lee, D. M. An investigation of longitudinal and transverse resonance in superfluid ^3He -A in low magnetic fields. *J. Low Temp. Phys.* **24**, 563–577 (1976).
14. Parpia, J. M. & Reppey, J. D. Critical velocities in superfluid ^3He . *Phys. Rev. Lett.* **43**, 1332–1336 (1979).
15. Freeman, M. R., Germain, R. S., Thuneberg, E. V. & Richardson, R. C. Size effects in thin films of superfluid ^3He . *Phys. Rev. Lett.* **60**, 596–599 (1988).
16. Monien, H. & Tewordt, L. Longitudinal magnetic resonance (NMR) in A_1 phase of superfluid ^3He . *J. Low Temp. Phys.* **60**, 323–333 (1985).
17. Schopohl, N., Marquardt, W. & Tewordt, L. Magnetic field dependence of ultrasound attenuation and sound velocity in superfluid ^3He -A. *J. Low Temp. Phys.* **59**, 469–485 (1985).
18. Einzel, D., Wölfle, P. & Hirschfeld, P. J. Transverse surface impedance of pair-correlated Fermi liquids. Application to normal and superfluid ^3He . *J. Low Temp. Phys.* **80**, 31–68 (1990).
19. Corruccini, L. R. & Osheroff, D. D. Unusual magnetic relaxation in superfluid ^3He . *Phys. Rev. Lett.* **34**, 564–567 (1975).
20. Osheroff, D. D. & Anderson, P. W. Nuclear magnetic resonance of superfluid ^3He near T_c in high magnetic fields. *Phys. Rev. Lett.* **33**, 686–689 (1974).
21. Bastea, M. et al. Spin entropy wave propagation in magnetically polarized superfluid ^3He . *J. Low Temp. Phys.* **137**, 539–578 (2004).
22. Roobol, L. P. et al. Shear viscosity of superfluid ^3He up to 15 tesla. *Phys. Rev. Lett.* **79**, 685–688 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Masutomi and K. Kimura for their contributions in the early stages of the experiments and A. Leggett for correspondence. This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas of MEXT Japan and JPSJ, and by the Condensed Matter Physics and East Asia and Pacific Programs of the US National Science Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.K. (kojima@physics.rutgers.edu).

Patterning organic single-crystal transistor arrays

Alejandro L. Briseno^{1,2}, Stefan C. B. Mannsfeld¹, Mang M. Ling¹, Shuhong Liu¹, Ricky J. Tseng², Colin Reese¹, Mark E. Roberts¹, Yang Yang², Fred Wudl² & Zhenan Bao¹

Field-effect transistors made of organic single crystals are ideal for studying the charge transport characteristics of organic semiconductor materials¹. Their outstanding device performance^{2–8}, relative to that of transistors made of organic thin films, makes them also attractive candidates for electronic applications such as active matrix displays and sensor arrays. These applications require minimal cross-talk between neighbouring devices. In the case of thin film systems, simple patterning of the active semiconductor layer^{9,10} minimizes cross-talk. But when using organic single crystals, the only approach currently available for creating arrays of separate devices is manual selection and placing of individual crystals—a process prohibitive for producing devices at high density and with reasonable throughput. In contrast, inorganic crystals have been grown in extended arrays^{11–13}, and efficient and large-area fabrication of silicon crystalline islands with high mobilities for electronic applications has been reported^{14,15}. Here we describe a method for effectively fabricating large arrays of single crystals of a wide range of organic semiconductor materials directly onto transistor source–drain electrodes. We find that film domains of octadecyltriethoxysilane microcontact-printed onto either clean Si/SiO₂ surfaces or flexible plastic provide control over the nucleation of vapour-grown organic single crystals. This allows us to fabricate large arrays of high-performance organic single-crystal field-effect transistors with mobilities as high as 2.4 cm² V^{−1} s^{−1} and on/off ratios greater than 10⁷, and devices on flexible substrates that retain their performance after significant bending. These results suggest that our fabrication approach constitutes a promising step that might ultimately allow us to utilize high-performance organic single-crystal field-effect transistors for large-area electronics applications.

Figure 1a illustrates the single-crystal patterning process (see Methods for a detailed description). First, thick octadecyltriethoxysilane (OTS) films (average thickness $\sim 13 \pm 2$ nm, measured by ellipsometry) are printed by microcontact printing onto a clean SiO₂/Si substrate, using a polydimethylsiloxane (PDMS) stamp with a relief structure in the desired pattern. We note that although the results discussed here mainly relate to structures grown on arrays of micrometre-sized OTS squares, it is straightforward to create more-complex OTS film patterns for single-crystal nucleation (Supplementary Figs 3 and 6). After film deposition, crystals are grown using a vapour transport method¹⁶ applicable to a broad range of materials, including high-mobility p-type materials such as rubrene, pentacene and tetracene, and n-type materials such as C₆₀, fluorinated copper phthalocyanine (F₁₆CuPc) and tetracyanoquinodimethane (TCNQ) (Supplementary Fig. 14). We find that sublimation of the organic material and crystal growth are accomplished in as little as five minutes for pentacene and as much as two hours for C₆₀, and that crystal nucleation is restricted to OTS-stamped regions and not observed on the SiO₂ background. The crystals show strong birefringence in optical micrographs recorded under cross-polarized

light (Supplementary Fig. 6), confirming their crystalline nature. We also observe intense and narrow diffraction peaks in the X-ray diffraction patterns of pentacene, rubrene and C₆₀ single crystals, which are indicative of a high degree of crystallinity (Supplementary Figs 5, 8 and 12).

Examples of arrays of single crystals of pentacene, rubrene and C₆₀ grown on stamped OTS domains are shown in Fig. 1b–d. The growth of discrete single crystals is achieved by controlling the feature size of the printed OTS regions so that it supports on average only one crystal nucleation event. In the case of pentacene, crystal nucleation activity as a function of the size of the stamped OTS domains was analysed by using arrays of OTS squares with areas of 25 μm^2 (that is, $5 \times 5 \mu\text{m}$), 625 μm^2 ($25 \times 25 \mu\text{m}$), 2,500 μm^2 ($50 \times 50 \mu\text{m}$) and 10,000 μm^2 ($100 \times 100 \mu\text{m}$). To ensure that crystallization can be accurately compared on different active areas, a special PDMS stamp was fabricated containing four different feature sizes on one stamp, thus preventing sample-to-sample variations. The results are summarized in Fig. 2a, which plots the average number of nuclei (as counted in numerous fields of each stamp size) versus the stamped domain area. Within experimental error, the data fit a line corresponding to a constant nucleation density of about one nucleus per 49 μm^2 , or a single pentacene crystal per $7 \times 7 \mu\text{m}$ domain area. This finding is in good agreement with our finding that the $5 \times 5 \mu\text{m}$ stamp (used to generate the array shown in Fig. 1b) yields approximately one single crystal per domain. The images in Fig. 2b show the results of using a large-area array of $5 \times 5 \mu\text{m}$ stamped OTS domains for controlling pentacene growth, demonstrating that most patterns result in the formation of discrete pentacene single crystals.

OTS surface treatments have previously been used to pattern pentacene thin films on structured substrates. In the case of preferential pentacene thin film growth on OTS-treated gold surface structures¹⁷, selective film growth was attributed to different surface interactions and surface diffusivities. Patterning SiO₂ substrates with OTS and FDTS ((heptafluorotetrahydrodecyl)trichlorosilane) also results in patterned growth of pentacene films⁹: within a certain window of growth conditions, pentacene forms two-dimensional islands (islands with monolayer-stepped terraces) on the patterned OTS regions, but only three-dimensional islands (disconnected single grains) in the FDTS-treated regions. These locally different modes of pentacene film growth were attributed to different molecule–substrate interactions.

However, the present study investigates the patterning of discrete single crystals, rather than the formation and patterning of thin films^{18,19}, and the selectivity achieved in our systems is more likely to be due to the high roughness of the thick OTS-stamped film rather than to different molecule–substrate interactions arising from a contrast in surface chemistry. This view is supported by atomic force microscope images typically showing an r.m.s. roughness of $\sim 150 \text{ \AA}$ in the stamped OTS domains (Supplementary Fig. 1), and an r.m.s. roughness of $< 5 \text{ \AA}$ in regions of bare SiO₂. In control experiments

¹Department of Chemical Engineering, Stanford University, Stanford, California 94305, USA. ²Department of Chemistry and Biochemistry and Exotic Materials Institute, Department of Materials Science and Engineering, University of California–Los Angeles, Los Angeles, California 90095, USA.

using thick OTS films stamped onto a vapour-deposited smooth OTS monolayer background (r.m.s. roughness $<2\text{ \AA}$), pentacene crystals nucleated only in the patterned regions and with the same selectivity as seen with thick OTS film patterns stamped on bare SiO_2 (Supplementary Fig. 15). These observations agree with classical nucleation theory, which predicts that the higher surface energy associated with irregularities such as indentions, step edges or protrusions significantly lowers the barrier to heterogeneous nucleation, relative to that on a flat surface^{20,21}; they also agree with a modern phase field theory, which similarly predicts heterogeneous nucleation to favourably occur on rough surfaces^{22,23}. However, we note that besides the rough topology of the OTS-stamped domains, the use of a rather high substrate temperature is also important in achieving selective crystal growth. The substrate temperature, which is only $\sim 20^\circ\text{C}$ lower than that of the evaporation source, results in a higher thermal redesorption rate in the bare SiO_2 regions than in the adjacent rough OTS regions. Given the large number of molecules required to form a stable nucleus at high substrate temperatures^{9,24}, nucleation on the bare SiO_2 surface is effectively suppressed. Consequently, the nucleation of crystals is effectively suppressed in smooth regions, while the stamped rough OTS domains act as effective primary nucleation centres. This interpretation is supported by inspection of the underside of the crystal making direct contact with the OTS film, which revealed that the initial nucleus may have grown from the valley of the rough OTS film in a conformal fashion, with large parts of the bottom surface being planar and in close contact to the substrate (Supplementary Figs 17, 18).

The control over crystal nucleation and growth provided by the OTS patterns makes it possible to grow large arrays of crystals directly

onto transistor source–drain electrodes (see also Methods). This is illustrated in Fig. 3a, which shows rubrene single crystals patterned into a 14×14 transistor array. In most cases a few crystals are found between each pair of source and drain electrodes, but in some cases just one rubrene single crystal bridges the channel region (optical micrograph in Fig. 3b). One-step fabrication of large arrays of transistors with a single crystal in the channel region has until now been impossible to achieve with any of the reported methods of single-crystal field-effect transistor fabrication¹. More than 99% of the devices with at least one crystal bridging the drain and source electrodes exhibited a field effect, with the electrical characteristics of the highest-mobility rubrene device shown in Fig. 3b. Overall, measured mobility values range from 0.1 to $2.4\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$. This large variability may be partly attributed to the marked anisotropy of charge transport in rubrene⁷ and to differences in contact quality at the dielectric/semiconductor and electrode/semiconductor interfaces.

The crystals are able to make contact because the stamped OTS on both Au and SiO_2 consists of sparse pillars that do not fully coat the surface. Moreover, some of the OTS stamped on Au was found to evaporate on heating under conditions similar to those used during crystal growth (Supplementary Fig. 19). The growing crystal itself also seems to push some OTS pillars outside the stamping field, thereby leaving a visible outline of the peeled-off crystal on the substrate (Supplementary Fig. 20); this effect increases the contact area between the crystal and the dielectric and electrode surfaces. But the calculated mobilities may nevertheless underestimate the true mobilities, because it is still possible that only portions of the bottom surface of the crystal make contact with the dielectric surface. Nonetheless, random testing of 22 of the devices in the 14×14

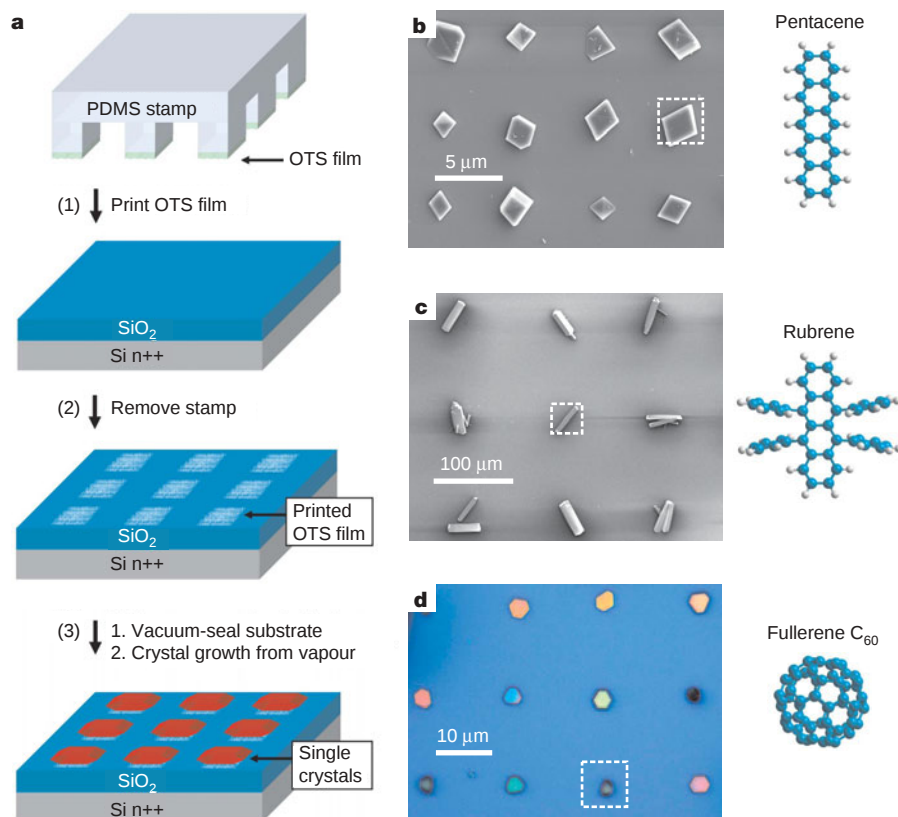


Figure 1 | Patterning of organic single crystals. **a**, Schematic outline of the procedure used to grow organic single crystals on substrates that have been patterned by microcontact printing. The printing uses PDMS stamps with relief features that are inked with a thick OTS film and then pressed onto the substrates. To grow the patterned single crystals, the patterned substrate is placed in a glass tube with the organic source material, vacuum-sealed (0.38 mm Hg), and placed in a temperature gradient furnace tube.

b–d, Patterned single-crystal arrays of different organic semiconductor materials. The dotted square in each image indicates the size and location of one of the OTS-stamped domains, while the molecular structure of the organic material used is shown next to the image of its single-crystal array. **b**, Pentacene (scanning electron microscope, SEM, image; stamped domain size $4 \times 4\text{ }\mu\text{m}$); **c**, rubrene (SEM, $25 \times 25\text{ }\mu\text{m}$); **d**, C_{60} (optical micrograph, $8 \times 8\text{ }\mu\text{m}$).

rubrene-based transistor array yielded an average saturation regime mobility of $\sim 0.6 \pm 0.5 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ with on/off ratios greater than 10^7 (Supplementary Fig. 10). These characteristics are comparable to those of amorphous silicon transistors and high-performance organic thin film transistors reported in the literature^{25–27}. Pentacene single-crystal transistors were similarly fabricated using unpurified, as-received pentacene source material, with the resultant devices exhibiting mobilities on the order of $0.3 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and on/off ratios $> 10^5$. The electrical characteristics of a typical pentacene device are displayed in Fig. 3c. Finally, devices made by patterning the n-channel materials C_{60} and TCNQ were found to exhibit mobilities of 0.03 and $10^{-4} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, respectively.

Our fabrication method can generate not only discrete single-crystal transistors on rigid SiO_2/Si substrates, but also flexible transistor arrays. The basic scheme for flexible single-crystal devices on a polyimide substrate and its realization with rubrene as the active material are shown in Fig. 4a (see also Methods for details). The characteristics

of a selected plastic rubrene transistor before and after bending (stretching, bending direction across the channel) are presented in Fig. 4b. Mobilities as high as $0.9 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and on/off ratios of 10^4 (threshold voltage $\sim 1.5 \text{ V}$) have been measured in our best flexible rubrene field-effect transistors, while our best mobility for a flexible pentacene field-effect transistor was $0.1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ with on/off ratios of $\sim 10^3$ (Supplementary Fig. 5). No significant loss in performance was observed when the devices were bent to a radius as small as 6 mm, as found in bending experiments performed on four randomly chosen devices that all exhibited similar behaviour. This demonstrates that the patterned single-crystal transistors can endure strenuous bending, and therefore are potentially useful for flexible electronics. Besides rubrene, high vapour pressure materials such as anthracene and tetracene were also successfully patterned on flexible and transparent poly(ethylene terephthalate) (PET) substrates without degrading the plastic (Fig. 4a).

The present results clearly show that the fabrication method we describe here can be used to control the nucleation location and nucleation density of molecular single crystals. But for the method to be technologically relevant, the orientation and alignment of the crystals also need to be controlled, because charge transport in single crystals can be highly anisotropic⁷. Preliminary results from our laboratory suggest that this is possible by using a combined nucleation/alignment layer. Another challenge is the need for a better understanding of the crystal growth mechanism in this system so as

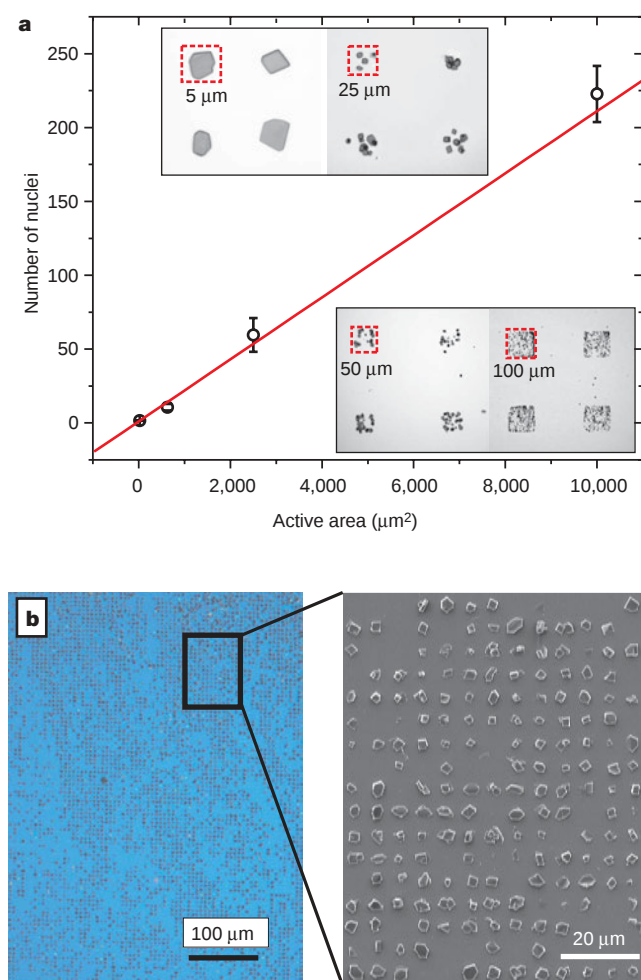


Figure 2 | Controlling pentacene crystal nucleation. The number of crystals nucleated and grown per stamped domain is a function of the domain area. **a**, Plot of the number of pentacene crystals in each stamped domain as a function of stamped domain area. The data points show the average crystal number found for the four different stamp sizes used in the experiment. The error bars indicate the standard error of the mean, obtained from counting the nuclei in 24 stamping domains of each size ($N = 24$). The red line is a linear fit of the experimental data, yielding a nucleation density value of one pentacene crystal per $49 \mu\text{m}^2$ (equivalent to a stamping area of $7 \times 7 \mu\text{m}$). Insets, optical micrographs of the four different stamp sizes used. **b**, Example of a large-area array of discrete pentacene single crystals grown on a $5 \times 5 \mu\text{m}$ stamped OTS pattern. The image on the left is an optical micrograph of the array, with the enlarged view of the indicated area shown as an SEM image on the right.

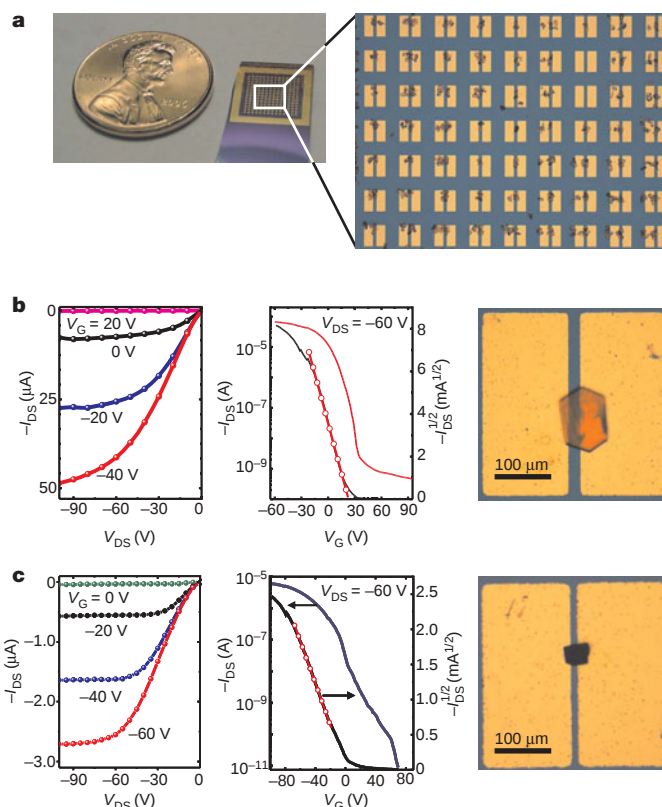


Figure 3 | Single-crystal transistor arrays on rigid substrates. Arrays of working single-crystal transistors are produced by printing films of OTS domains onto arrays of source-drain electrodes, followed by growth of organic single crystals in the printed domains. **a**, Optical micrograph of arrays of functional single-crystal transistors. The electrical characteristics of typical single-crystal transistors in the arrays, along with images of the transistors, are shown in **b** for rubrene, and in **c** for pentacene. V_{DS} , drain-source voltage; V_G , gate voltage; I_{DS} , drain-source current. Characteristics are as follows: **b** (middle panel), mobility (μ) = $2.4 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, on/off ratio $\sim 10^6$, threshold voltage (V_{th}) = 22.4 V ; **c** (middle panel), μ = $0.2 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, on/off ratio $\sim 10^6$, V_{th} = -3.1 V .

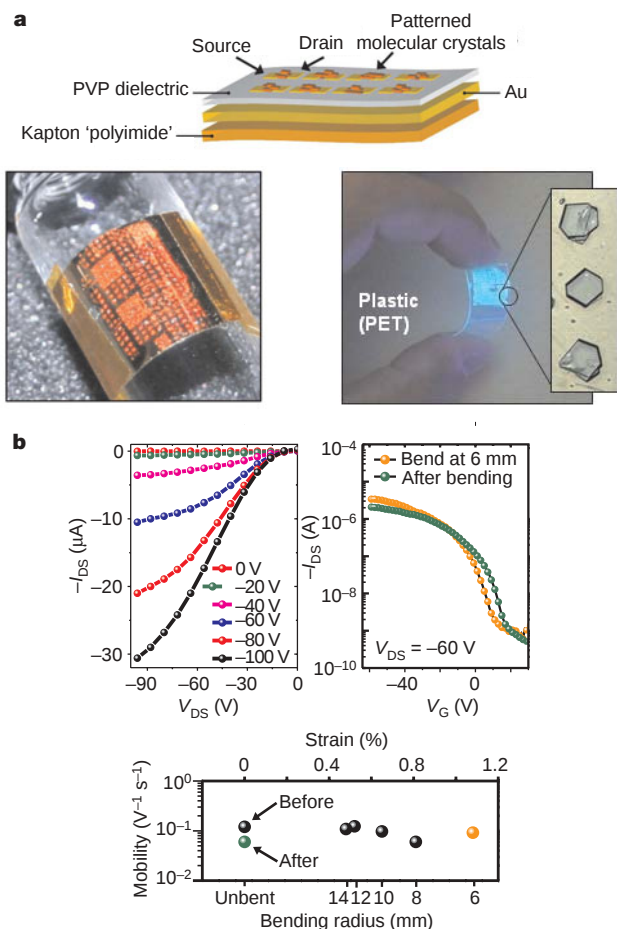


Figure 4 | Flexible single-crystal transistor arrays. **a**, Diagram of the basic design for single-crystal transistor arrays on plastic substrates (top). The optical micrographs (bottom) show a rubrene transistor array attached to a glass vial of diameter 11.7 mm (left), and anthracene crystals patterned on a poly(ethylene terephthalate) (PET) substrate and fluorescing under an ultraviolet light source (right). Patterned organic crystals were grown on flexible Kapton devices that were taped onto pre-cut glass microscope slides with high-temperature polyimide tape. **b**, Electrical characteristics of the single-crystal arrays. Left, the performance of a typical flexible rubrene single-crystal transistor. The middle and right panels illustrate the effect of device array bending on device performance. Middle, the transfer curves of a flexible device bent at 6 mm bending radius and after bending (that is, back in planar geometry); right, plot of mobility as a function of the bending radius (or strain percentage, which is indicated on the top axis). The green and orange curves in the middle panel correspond to the green and orange data points in the right panel. The electrical characteristics of all devices were first measured in the planar geometry, then repeatedly in the bent geometry, and finally again in the original planar geometry, similar to procedures used in ref. 28.

to optimize the electrical contact between the crystals and the dielectric layer, as well as between the crystals and the electrodes. Of course, practical applications in large-area flexible electronics will also require precise control of the growth process, to allow scaling of the fabrication method to larger areas while achieving high device uniformity and high device yield.

METHODS

Substrate patterning and crystal growth. OTS printing solutions were prepared in reagent grade toluene with concentrations ranging from 0.1 to 1.0 M. Cotton applicators were used to ink the PDMS stamps with the OTS solutions, and an air gun was used to dry the stamps from excess solvent. The stamps were brought into contact with substrates for approximately 5 min. No rinsing or heat treatment was administered to the OTS-patterned substrates. To deposit the crystals, we used a 16-mm-diameter glass tube that was open at one end and had already

been closed (by flame-sealing) at the other. Semiconductor material was placed inside the tube near the closed end (~ 10 mg of source material used); the patterned substrate was put onto a NiCr wire clip and placed inside the tube, several centimetres away from the source material. Subsequently, the tube was completely flame-sealed under a vacuum pressure of ~ 0.38 mm Hg using a standard laboratory vacuum pump. As an example of the procedure used, the pentacene source temperature is maintained at $\sim 260^\circ\text{C}$ and the closest edge of the patterned substrate is positioned 2 cm away ($\sim 250^\circ\text{C}$) from the source zone. The nucleation zone of pentacene crystals extends to the far edge of the sample 5 cm away from the source ($\sim 220^\circ\text{C}$). Next, the vacuum-sealed tube is placed into a temperature gradient sublimation reactor similar to the construction reported in ref. 16. Pentacene single crystals were grown in as little as 5 min with $5\text{ }\mu\text{m}$ stamps and in ~ 30 min using $100\text{ }\mu\text{m}$ stamps. Typical source temperatures for pentacene are $\sim 260^\circ\text{C}$. Growth temperatures for rubrene and C_{60} were $250\text{--}270^\circ\text{C}$ and $425\text{--}500^\circ\text{C}$, respectively. Increasing the source temperature increases the rate of crystal growth, but also increases the distance from the source to the crystal-forming region. Calibration of materials under controlled parameters was conducted to determine the exact distance of crystal growth with respect to the source zone. Larger-diameter glass tubes were used in order to fabricate larger-area patterns. Organic source materials were purified by temperature gradient sublimation, but in most cases, they were used without any purification and used as received.

Device array fabrication. Bottom-contact devices (14×14 source-drain electrode arrays, channel lengths 10, 20, 50, $100\text{ }\mu\text{m}$) were defined by conventional photolithography (metal thickness: Cr, 1.5 nm; Au, 40 nm). Single-crystal transistors were fabricated on conventional SiO_2 substrates containing a dielectric thickness of 300 nm on highly doped Si. A capacitance of 10 nF cm^{-2} was used to calculate the field-effect mobility of transistors on a 300 nm SiO_2 dielectric, whereas a capacitance of 1.9 nF cm^{-2} was used for $1.5\text{-}\mu\text{m}$ -thick poly-4-vinyl-phenol (PVP, $M_w = 20,000$) dielectric. For flexible transistors, a dielectric layer of PVP was spin-coated ($\sim 2,000$ r.p.m.) on $140\text{-}\mu\text{m}$ -thick Kapton (polyimide) sheets covered with a 100 nm gold coating used for the gate electrode (Astral Technology Unlimited). The dielectric solution was prepared from 22 wt% PVP, and 8 wt% poly(melamine-co-formaldehyde), methylated ($M_w = 511$), dissolved in propylene glycol monomethyl ether acetate (PGMEA)²⁷. The substrates were baked at 100°C for 10 min then at 200°C for 10 min to initiate crosslinking. The source-drain electrodes were fabricated by thermal evaporation of Cr (1.5 nm) and Au (50 nm) for flexible substrates. The W/L (W , channel width; L , channel length) for our highest mobility rubrene devices were ~ 2 for conventional SiO_2 substrates and ~ 4 for flexible polyimide devices. Patterned organic crystals were grown on flexible Kapton devices that were taped onto pre-cut glass microscope slides with high-temperature polyimide tape. All devices were first measured in the planar geometry mode followed by several bent geometry measurements, and finally measured in the original planar geometry, similar to procedures used in ref. 28.

Received 6 July; accepted 1 November 2006.

- de Boer, R. W. I., Gershenson, M. E., Morpurgo, A. F. & Podzorov, V. Organic single-crystal field-effect transistors. *Phys. Status Solidi a* **201**, 1302–1331 (2004).
- Briseno, A. L. *et al.* Patterned growth of large oriented organic semiconductor single crystals on self-assembled monolayer templates. *J. Am. Chem. Soc.* **127**, 12164–12165 (2005).
- Huitema, H. E. A. *et al.* Plastic transistors in active-matrix displays. *Nature* **414**, 599 (2001).
- Moon, H. *et al.* Synthesis, crystal structure, and transistor performance of tetracene derivatives. *J. Am. Chem. Soc.* **126**, 15322–15323 (2004).
- Podzorov, V. *et al.* Intrinsic charge transport on the surface of organic semiconductors. *Phys. Rev. Lett.* **93**, 086602 (2004).
- Pope, M., Charles, E. & Swenberg, C. E. *Electronic Processes in Organic Crystals and Polymers* (Oxford Univ. Press, New York, 1999).
- Sundar, V. C. *et al.* Elastomeric transistor stamps: Reversible probing of charge transport in organic crystals. *Science* **303**, 1644–1646 (2004).
- Zeis, R. *et al.* Field effect studies on rubrene and impurities of rubrene. *Chem. Mater.* **18**, 244–248 (2006).
- Steudel, S., Janssen, D., Verlaak, S., Genoe, J. & Heremans, P. Patterned growth of pentacene. *Appl. Phys. Lett.* **85**, 5550–5552 (2004).
- De Vusser, S., Steudel, S., Myny, K., Genoe, J. & Heremans, P. Integrated shadow mask method for patterning small molecule organic semiconductors. *Appl. Phys. Lett.* **88**, 103501 (2006).
- Aizenberg, J., Black, A. J. & Whitesides, G. M. Controlling local disorder in self-assembled monolayers by patterning the topography of their metallic supports. *Nature* **394**, 868–871 (1998).
- Aizenberg, J., Black, A. J. & Whitesides, G. M. Control of crystal nucleation by patterned self-assembled monolayers. *Nature* **398**, 495–498 (1999).
- He, J. H. *et al.* Pattern and feature designed growth of ZnO nanowire arrays for vertical devices. *J. Phys. Chem. B* **110**, 50–53 (2006).

14. van der Wilt, P. C., van Dijk, B. D., Bertens, G. J., Ishihara, R. & Beenakker, C. I. M. Formation of location-controlled crystalline islands using substrate-embedded seeds in excimer-laser crystallization of silicon films. *Appl. Phys. Lett.* **79**, 1819–1821 (2001).
15. Vikas, R. *et al.* in *Proc. International Electron Devices Meeting 2005* 919–922 (IEEE International, 2005).
16. Kloc, C., Simpkins, P. G., Siegrist, T. & Laudise, R. A. Physical vapor growth of centimeter-sized crystals of alpha-hexathiophene. *J. Cryst. Growth* **182**, 416–427 (1997).
17. Choi, H. Y., Kim, S. H. & Jang, J. Self-organized organic thin-film transistors on plastic. *Adv. Mater.* **16**, 732–736 (2004).
18. Ruiz, R. *et al.* Pentacene thin film growth. *Chem. Mater.* **16**, 4497–4508 (2004).
19. Heringdorf, F., Reuter, M. C. & Tromp, R. M. Growth dynamics of pentacene thin films. *Nature* **412**, 517–520 (2001).
20. Markov, I. V. *Crystal Growth for Beginners* (World Scientific, Singapore, 2003).
21. Venables, J. A., Spiller, G. D. T. & Hanbucken, M. Nucleation and growth of thin films. *Rep. Prog. Phys.* **47**, 399–459 (1984).
22. Granasy, L., Borzsonyi, T. & Pusztai, T. Nucleation and bulk crystallization in binary phase field theory. *Phys. Rev. Lett.* **88**, 206105 (2002).
23. Granasy, L. *et al.* Phase field theory of crystal nucleation and polycrystalline growth: A review. *J. Mater. Res.* **21**, 309–319 (2006).
24. Verlaak, S., Steudel, S., Heremans, P., Janssen, D. & Deleuze, M. S. Nucleation of organic semiconductors on inert substrates. *Phys. Rev. B* **68**, 195409 (2003).
25. Kelley, T. W. *et al.* Recent progress in organic electronics: Materials, devices, and processes. *Chem. Mater.* **16**, 4413–4422 (2004).
26. Klauk, H., Gundlach, D. J., Nichols, J. A. & Jackson, T. N. Pentacene organic thin-film transistors for circuit and display applications. *IEEE Trans. Electron Devices* **46**, 1258–1263 (1999).
27. Klauk, H. *et al.* High-mobility polymer gate dielectric pentacene thin film transistors. *J. Appl. Phys.* **92**, 5259–5263 (2002).
28. Briseno, A. L. *et al.* High-performance organic single-crystal transistors on flexible substrates. *Adv. Mater.* **18**, 2320–2324 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements A.L.B. acknowledges a Bell Labs Graduate Research Fellowship. A.L.B., F.W., R.J.T. and Y.Y. acknowledge financial support from the Air Force Office of Scientific Research (AFOSR). S.C.B.M. acknowledges financial support from the Deutsche Forschungsgemeinschaft, and Z.B. acknowledges partial support from the Stanford Center for Polymeric Interfaces and Macromolecular Assemblies (NSF-Center MRSEC), the Stanford School of Engineering, and a Sloan Research Fellowship.

Author Contributions A.L.B. performed single-crystal patterning on substrates and FET devices, measurements on FET devices, and wrote most of the manuscript. S.C.B.M. elucidated the growth mechanism, performed AFM measurements, and wrote parts of the manuscript. M.M.L. assisted in sample preparation and assisted in device measurements and calculations. S.L. assisted in crystal patterning, device preparation, and performed SEM measurements. R.J.T. assisted in flexible device preparation, measurements and calculations. C.R. designed and fabricated the transistor array devices and PDMS stamp masters via lithography. M.E.R. assisted in early experiments and purified some of the organic source materials. Y.Y. provided use of laboratory space and instruments. F.W. provided feedback and suggestions on experiments. Z.B. conceptualized and directed the research project. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Z.B. (zbao@chemeng.stanford.edu).

LETTERS

Negligible glacial–interglacial variation in continental chemical weathering rates

Gavin L. Foster¹ & Derek Vance¹

Chemical weathering of the continents is central to the regulation of atmospheric carbon dioxide concentrations, and hence global climate^{1,2}. On million-year timescales silicate weathering leads to the draw-down of carbon dioxide¹, and on millennial timescales chemical weathering affects the calcium carbonate saturation state of the oceans and hence their uptake of carbon dioxide². However, variations in chemical weathering rates over glacial–interglacial cycles remain uncertain. During glacial periods, cold and dry conditions reduce the rate of chemical weathering³, but intense physical weathering^{3,4} and the exposure of carbonates on continental shelves due to low sea levels^{5,6} may increase this rate. Here we present high-resolution records of the lead isotope composition of ferromanganese crusts from the North Atlantic Ocean that cover the past 550,000 years. Combining these records with a simple quantitative model of changes in the lead isotope composition of the deep North Atlantic Ocean in response to chemical weathering, we find that chemical weathering rates were two to three times lower in the glaciated interior of the North Atlantic Region during glacial periods than during the intervening interglacial periods. This decrease roughly balances the increase in chemical weathering caused by the exposure of continental shelves, indicating that chemical weathering rates remained relatively constant on glacial–interglacial timescales. On timescales of more than a million years, however, we suggest that enhanced weathering of silicate glacial sediments during interglacial periods results in a net draw-down of atmospheric carbon dioxide, creating a positive feedback on global climate that, once initiated, promotes cooling and further glaciation.

The oceans integrate Pb released by weathering from relatively large continental areas⁷, and records of secular change in the Pb isotopic composition of seawater yield such information on long timescales. Ferromanganese crusts (hereafter simply called crusts) precipitate directly from seawater and have been shown by numerous publications to provide a faithful record of its Nd and Pb isotopic composition⁸. We focus here on three crusts that have been well dated with U-series (Supplementary Information): ALV539 2-1 (from here on termed ALV539) and BM1969.05 from the New England and adjacent San Pablo Seamounts in the northwest Atlantic (35° 36' N 58° 47' W, 2,665 m; 39° 00' N 60° 57' W, 1,800 m), and TR079 D-14 from the Desirade Seamount in the Guadeloupe Islands of the Lesser Antilles in the tropical west Atlantic (16° 55' N 61° 10' W, about 2,000 m). All three are currently bathed by the Deep Western Boundary Current, which is the southern extension of North Atlantic Deep Water (NADW). At present the NADW is derived from a roughly 70%/30% mixture of surface waters in the Greenland–Iceland–Norwegian Seas and in the Labrador Sea⁹.

Figure 1 shows a new high-resolution Pb isotope record for these crusts, obtained with laser-ablation multi-collector inductively coupled plasma mass spectrometry (LA-MC-ICPMS; Supplementary

Methods). Four asymmetric cycles are evident in the isotopes of Pb in the outer 550 kyr of ALV539 and BM1969.05 (Fig. 1) that clearly correlate with the O isotope record of seawater despite subtle variations (typically on timescales of the order of less than 15 kyr) relating to uncertainties in the crust age models used (principally due to small-scale variations in growth rate; Supplementary Information). The Pb isotope record shows asymmetric roughly 100-kyr cycles, with ²⁰⁷Pb/²⁰⁶Pb varying from 0.824 to 0.817 during the glacial stages and increasing rapidly to about 0.814 during interglacials. All other Pb isotope ratios (for example ²⁰⁶Pb/²⁰⁴Pb) show similar periodicities (Supplementary Information). Significantly for our interpretation, crust TR079 D-14 from 16° N has much less variability.

Secular variation in the Pb isotopic composition of the deep North Atlantic, as recorded in crusts, could reflect either variable mixing between different water masses in response to a change in the pattern of deep ocean circulation or changes in the rate and mode of weathering on the North Atlantic continents (see ref. 8 for a review). There are several lines of evidence that the latter dominates. First, Pb has a very short residence in the North Atlantic (less than 30 years (ref. 7)) and therefore Pb is not advected far by Atlantic Ocean circulation (water has a residence time of about 600 years in the Atlantic). That the Pb isotopic signal in the northern crusts is dominated by local weathering processes and is not affected by mixing of water from the south is supported by the lack of variation in the southernmost crust, TR079 D-14. Second, tracers of water mass such as Nd isotopes¹⁰ (Fig. 1) indicate near-constant modern-NADW-like¹¹ isotope ratios in the crusts and show that there has been little impact of southern-derived water for the past 550 kyr (see ref. 10) (Fig. 1). Last, Fig. 2a shows that over the past 2.2 Myr the NADW has been significantly more radiogenic in Pb than its potential source areas. It has been established in numerous areas that the early stages of weathering preferentially release radiogenic Pb (for example ²⁰⁷Pb/²⁰⁶Pb as low as 0.680 in soils versus 0.957 in the parent rock¹²) from high-U + Th/Pb accessory phases (Fig. 2b)^{12–14}. It has been suggested¹⁵ that the radiogenic Pb at one end of the arrays in Fig. 2 is released as a consequence of this process and that it was significantly augmented after the beginning of Northern Hemisphere glaciation (2.5–3.0 Myr ago) as a result of the increased mechanical breakdown of accessory grains by glacial erosion. Although Nd isotopes are also subject to the impact of incongruent weathering, the effects are more subtle¹⁵. Moreover, the exotic Nd produced by incongruent weathering is released into a much larger oceanic reservoir of Nd (a consequence of its longer residence time), so that the impact on seawater is more subtle still. The data in Fig. 1 do not preclude the presence of such small excursions in seawater Nd as sampled by the crusts.

We therefore argue that the dominant process causing variation in ALV539 and BM1969.05 up and down the arrays in Fig. 2 and the glacial–interglacial variation in Pb isotope ratio is the differential release of radiogenic Pb from accessory phases during the early stages

¹Bristol Isotope Group, Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queens Road, Bristol BS8 1RJ, UK.

of weathering. Figure 2c can be interpreted to yield ages of 2.0–2.5 Gyr for the source, very similar to Nd model ages for surface seawater from the Labrador Sea over the past 1 Myr^{16,17} and also to Ar–Ar ages of detrital hornblendes in North Atlantic sediments¹⁷ and Pb isotope data (Fig. 2b, c) for leachates of Baffin Bay sediments¹⁵. This result could only coincidentally be a product of water mass mixing.

The lack of Pb isotopic cycles in TR079 D-14 indicates that the origin of the signal must be quite proximal to ALV539 and BM1969.05; that is, originating on the North American continent. It has been shown¹² that the radiogenic fraction of Pb released during early chemical weathering of soils is soon swamped by the release of Pb from other less radiogenic phases such that after 300 kyr the Pb isotopic composition of the soil is indistinguishable from the unweathered parent rock (Fig. 2b). Over a similar timescale (about 100 kyr) the chemical weathering rate of a soil decreases by three

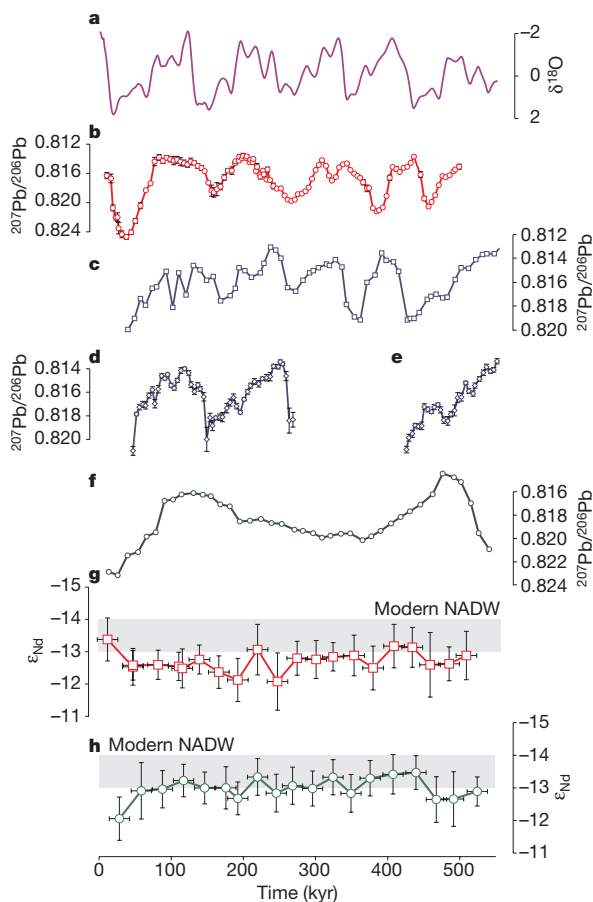


Figure 1 | High-resolution Pb and Nd isotope data for deep Atlantic ferromanganese crusts. **a**, The SPECMAP marine oxygen isotope record²⁷. **b–h**, High-resolution LA-MC-ICPMS Pb (**b–f**) and Nd (ref. 12) (**g, h**) isotope records of BM1969.05 (**b, g**), ALV539 A (**c**), ALV539 B (**d**), ALV539 C (**e**) and TR079 D-14 (**f, h**) plotted against time from U-series isotope analysis (Supplementary Methods, and ref. 10). The ϵ notation used in **g** and **h** is in parts per 10,000 variation from the Chondritic Uniform Reservoir. Nd and Pb error bars are 2 s.e. measurement uncertainties and for Pb they are typically smaller than the symbols. See Supplementary Information for the location of each Pb isotope traverse, for the tabulated data and plots of all Pb isotope ratios. For crusts BM1969 and TR079 D-14, as well as the youngest roughly 200 kyr of ALV539, age uncertainties are 10–15 kyr (Supplementary Methods). These age uncertainties do not significantly compromise the assignment of a part of the Pb isotope record to a glacial–interglacial. For the older part of the ALV539 record the age uncertainties increase and are as great as ± 50 kyr. However, the similarity of the ALV539 Pb isotope record to that of the better-dated BM1969.05 significantly increases confidence in the assignment of ages to the older parts of this crust.

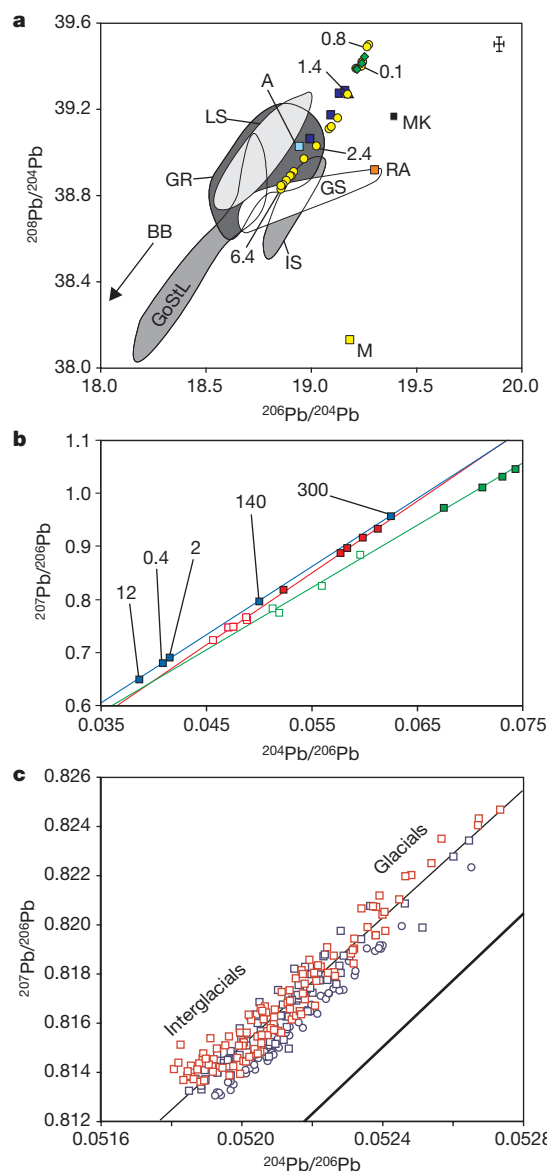


Figure 2 | Relationships between ferromanganese crust Pb isotopic data and circum-Atlantic Pb sources. **a**, Low-resolution²⁴ (labelled with ages in Myr), and some new high-resolution Pb isotope data, for ALV539 plotted with potential Pb sources (shaded fields). Source data are for either riverine suspended load²⁷ or geographically constrained oceanic sediments^{28,29}. Data for the NADW over the past 2.2 Myr clearly extend to values more radiogenic than any known source. Error bars (top right) on the high-resolution data are the external reproducibility of NIST 612 by LA-MC-ICPMS (Supplementary Methods). A, Amazon river; LS, Labrador Rise; GR, Greenland Rise; GS, Greenland Shelf; BB, Baffin Bay; RA, Red Arctic river; MK, MacKenzie river; IS, Iceland Shelf; GoStL, Gulf of St Lawrence; M, Mississippi river. Blue squares, ALV539 glacials (0–450 kyr; LA-MC-ICPMS); green diamonds, ALV539 interglacials (0–450 kyr; LA-MC-ICPMS); yellow circles, ALV539 (0–6 Myr; ref. 24). **b**, $^{207}\text{Pb}/^{206}\text{Pb}$ data for soils (Wind river¹²) labelled with soil age in kyr (blue), Baffin Bay (red) and Greenland river (green) sediments¹⁵ (leached (open red and blue symbols) and bulk (filled red and blue symbols) dissolutions). The Pb released by leaching of sediments is as enriched in radiogenic Pb as it is in young soils. **c**, Plot of $^{207}\text{Pb}/^{206}\text{Pb}$ against $^{204}\text{Pb}/^{206}\text{Pb}$ for BM1969.05 (red squares), ALV539 A (blue circles) and ALV539 B and C (blue squares). The $^{207}\text{Pb}/^{206}\text{Pb}$ versus $^{204}\text{Pb}/^{206}\text{Pb}$ data for both samples fall on statistically reasonable lines ($R^2 > 0.9$) whose intercepts give ages from 2.0 to 2.5 Gyr, in approximate agreement with model ages for large parts of the Canadian, Greenland and Scandinavian Shields. Also shown is the best-fit line (thick line) for leached Baffin Bay sediments (1.7 ± 0.4 Gyr)¹⁵. The similarity in the generated age provides support for the suggestion that crust arrays are produced by differential weathering (leaching) during soil formation.

orders of magnitude⁴, predominantly as a result of the decline in availability of fresh and easily weatherable mineral surfaces¹⁸. The isotopic contrast between the labile Pb and its parent rock Pb is therefore inversely proportional to the age of the soil; the isotopic contrast is negligible for soil more than 300 kyr old and is proportional to the rate of chemical weathering of the soil. It has been strongly argued¹⁸ that the processes observed in soils that developed in both old shield rocks in Wyoming and Mesozoic granitoids in the Sierra Nevada¹² are representative of glaciated silicate shield terrains. The glaciated terrains of the North American continent are dominated by such lithologies.

The extensive availability of fresh soil-parent material left behind by retreating continental ice sheets, in addition to the large amounts of melt water and warmer temperatures, would have caused chemical weathering rates to be highest during the early stages of interglacials¹⁸. This fresh soil-parent material, whether truly freshly generated or only freshly exposed to the agents of chemical weathering, would have yielded Pb with a relatively radiogenic isotopic composition¹². As the soils aged over the interglacial the amount of weathering would have decreased⁴, and there would have been a gradual decline in the radiogenic nature of the Pb released¹². The output of a simple quantitative model (see Methods), showing the changes in the Pb isotope composition of the deep ocean that would have resulted from such a process, is compared with the measured data in Fig. 3.

This model, which successfully describes our Pb data (Fig. 3), also predicts the average age of soils in the glaciated regions of the North American continent and can therefore give some insights into the variability in chemical weathering rate by using a published relationship between soil age and cationic weathering flux⁴. The average weathering rate the model predicts for marine isotope stage 1 (0–11 kyr ago) is 32 mequiv m⁻² yr⁻¹, which, for example, is similar to the modern average rate of about 37 mequiv m⁻² yr⁻¹ of the Slave province of Canada (calculated from ref. 19). The model allows us to estimate that, on average (excluding stage 1), 2.5 ± 0.6-fold more weathering occurs during warm interglacial periods than during glacial periods. This factor applies only to those regions that are glaciated or covered in periglacial sediment, which has been estimated to be about 25% of the continental landmass²⁰. Scaled according to land area, our data indicate that globally, as a result of the abundance of

fresh soil-parent material, the average chemical weathering rate during interglacials may be about 20–50% higher than during glacials. Much higher rates are indicated during the very early stages of an interglacial (see Fig. 3). This estimate of the change in average chemical weathering rate is similar in magnitude to that permissible by the Sr isotope composition of marine carbonates²¹. Such a decrease in the flux of cations from the continents to the oceans during glacial times will be partly balanced by the exposure, by a fall in the sea level, of easily weatherable shelf carbonates, which is thought to enhance the cation weathering flux by about 25%^{5,6}. The similar magnitude of these perturbations may serve to maintain a balance in the flux of cations to the oceans on short to intermediate timescales (10³–10⁶ years), providing an explanation for the lack of significant change in the calcium carbonate compensation depth in the world's oceans during glacial times²² despite changes in the nature and location of chemical weathering²⁰ and calcium carbonate production²². However, the rapid increase in weathering rate and consequent flux of cations to the oceans after deglaciation may not be adequately balanced and provides a plausible explanation for the ocean-wide carbonate preservation event (a result of increased ocean alkalinity) centred on 9 kyr (ref. 23).

On longer timescales, and in the context of our model of oceanic Pb isotopes, the gradual shift to more radiogenic Pb values that occurred in both ALV539 and BM1969.05 about 2.5 Myr ago²⁴ (Fig. 2a) probably relates to an overall decrease in the average soil age in the glaciated regions of North America as glacial erosion gradually removed the preglacial regolith²⁵. Preliminary *in situ* Pb data from these same crusts for the period before that highlighted here indicate that short-term high-amplitude changes in Pb isotopes are lacking before the mid-Pleistocene epoch. The gradual replacement of weathered Jurassic saprolites that previously blanketed the North American continent²⁵ by fresher material would have led to greater short-term variability in marine Pb isotopes and an overall increase in weathering rates¹⁸.

It has been suggested that weathering rates approach a steady-state value of 1–5 mequiv m⁻² yr⁻¹ for soils older than about 150 kyr, a value that may be representative of rates before the onset of the Northern Hemisphere glaciation. The average weathering rate predicted by our model for the past 550 kyr is about 9 mequiv m⁻² yr⁻¹, or about three times this value. As silicates predominantly underlie the Laurentide ice sheet⁵ and the chemical weathering of silicates results in a net draw-down of CO₂, this enhanced silicate weathering, which occurs predominantly during warm interglacial periods, would increase the long-term draw-down of CO₂ and contribute to global cooling, acting as a strong positive feedback serving to promote further glaciation.

METHODS

Quantitative model of Pb isotopes in the deep North Atlantic. In the following the 'glaciated region' refers to those parts of the North American continent that, in the 550-kyr period modelled, are at some stage covered by ice. 'Non-glaciated' regions are those parts of the North American continent that have never been covered by ice. The proportion of the glaciated region covered by ice was assumed to vary linearly with $\delta^{18}\text{O}$ (ref. 26) between values of 0 (none of the glaciated region covered in ice) and 1 (all of it covered in ice) and the retreat and advance of ice sheets was simulated at 1-kyr resolution for the past 550 kyr. A linear relationship between $\delta^{18}\text{O}$ (which is predominantly ice volume) and the area covered by an ice cap is clearly an oversimplification but such a consideration is of only minor importance here. For each 1-kyr interval, the total area of glaciated region in front of the ice sheet at that time is split into age bins spaced at 1 kyr, corresponding to the time elapsed since it was last covered by ice.

The ²⁰⁷Pb/²⁰⁶Pb isotope ratio of pre-Quaternary Pb in the deep North Atlantic is about 0.825 (ref. 24). This presumably reflects the Pb released from the weathering of very old preglacial saprolites and thus the average isotopic composition of bulk rock Pb in the North Atlantic region. We have used this value of ²⁰⁷Pb/²⁰⁶Pb for 300-kyr-old soils, a value of 0.807 for the isotopic composition of the Pb released by zero-age soils¹² and a linear relationship between soil age and the ²⁰⁷Pb/²⁰⁶Pb ratio of weathered Pb between these two extremes¹². The integrated isotopic composition of the entire area of the glaciated region in front

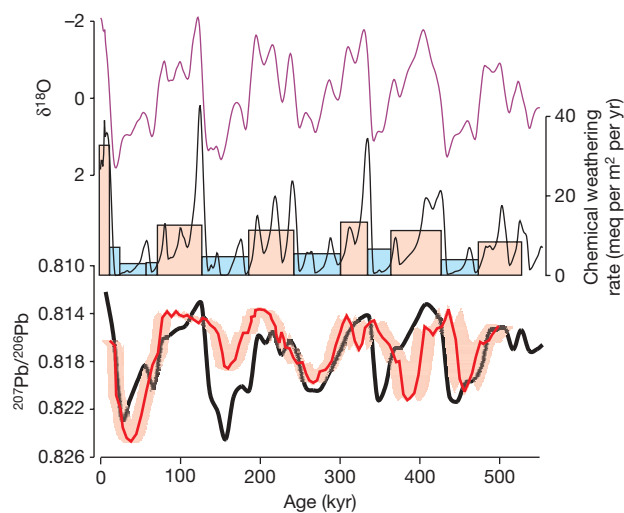


Figure 3 | Comparison between Pb isotopic data and output from a model based on variable glacial-interglacial weathering rates. Model Pb output (bottom, black line) and measured ²⁰⁷Pb/²⁰⁶Pb of BM1969.05 (red line) over the past 550 kyr with an error envelope relating to the uncertainty in the associated age model (Supplementary Methods). For the model weathering rate output (middle), bars are averages of the weathering rate during glacial (blue) and interglacial (red) marine isotope stages. The SPECMAP oxygen isotope stack is shown at the top²⁶. See Methods for a more detailed description of the model.

of the ice sheet was then calculated for each 1 kyr interval. No variation in the amount of Pb released as a function of soil age was modelled, because this is not merited by the amounts of leachable Pb in soils of various ages¹². The total flux of Pb, then, from the glaciated regions at any one time is purely a function of the fraction of the glaciated region exposed.

The processes described in ref. 18 and modelled above, operating on their own, result in much larger model glacial–interglacial swings in Pb than those observed, so we added a damping background Pb flux to the model. In the real world this is represented by the flux of Pb from non-glaciated regions to the deep North Atlantic. The quality of fit of the model to the real data is relatively insensitive to the magnitude of this flux, but the best fit is obtained when the flux from this source is about 0.3 times the flux from the glaciated region during an interglacial. The quality of fit of the model to the real data is much more sensitive to the isotopic composition of the background Pb flux and we gain some confidence from the fact that the best fit between model and data is obtained when its ²⁰⁷Pb/²⁰⁶Pb is the same as pre-Quaternary Pb in the deep North Atlantic, as recorded in long-term Fe–Mn crust records²⁴.

Although the fit of the resultant model output to the data is not perfect (for example, the depth of the Pb isotopic minima for marine isotope stage 6 and the less than 40-kyr discrepancy at stage 10/11), it does reproduce the major features of the data well and simulates the lows in ²⁰⁷Pb/²⁰⁶Pb associated with higher weathering rates and release of radiogenic Pb during interglacials.

Received 22 May; accepted 23 October 2006.

- Raymo, M. E., Ruddiman, W. F. & Froelich, P. N. Influence of late Cenozoic mountain building on ocean geochemical cycles. *Geology* **16**, 649–653 (1988).
- Archer, D., Winguth, A., Lea, D. W. & Mahowald, N. What caused the glacial/interglacial atmospheric pCO₂ cycles? *Rev. Geophys.* **38**, 159–189 (2000).
- Riebe, C. S., Kirchner, J. W. & Finkel, R. C. Erosional and climatic effects on long-term chemical weathering rates in granitic landscapes spanning diverse climate regimes. *Earth Planet. Sci. Lett.* **224**, 547–562 (2004).
- Taylor, A. & Blum, J. D. Relation between soil age and silicate weathering rates determined from chemical evolution of a glacial chronosequence. *Geology* **23**, 979–982 (1995).
- Gibbs, M. T. & Kump, L. R. Global chemical erosion during the last glacial maximum and the present: Sensitivity to changes in lithology and hydrology. *Paleoceanography* **9**, 529–543 (1994).
- Jones, I. W., Munhoven, G., Tranter, M., Huybrechts, P. & Sharp, M. J. Modelled glacial and non-glacial HCO₃[–], Si and Ge fluxes since the LGM: little potential for impact on atmospheric CO₂ concentrations and a potential proxy of continental chemical erosion, the marine Ge/Si ratio. *Global Planet. Change* **33**, 139–153 (2002).
- Henderson, G. M. & Maier-Reimer, E. Advection and removal of ²¹⁰Pb and stable Pb isotopes in the oceans: A general circulation model study. *Geochim. Cosmochim. Acta* **66**, 257–272 (2002).
- Frank, M. Radiogenic isotopes: tracers of past ocean circulation and erosional input. *Rev. Geophys.* **40**, doi:10.1029/2000RG000094 (2002).
- Dickson, R. R. & Brown, J. The production of North Atlantic Deep Water: Sources, rates and pathways. *J. Geophys. Res.* **99**, 12319–12341 (1994).
- Foster, G. L., Vance, D. & Prytulak, J. No change in the neodymium isotope composition of deep water exported from the North Atlantic on glacial–interglacial timescales. *Geology* (in the press).
- Piegras, D. J. & Wasserburg, G. J. Rare-earth element transport in the Western North-Atlantic inferred from Nd isotopic observations. *Geochim. Cosmochim. Acta* **51**, 1257–1271 (1987).
- Harlavan, Y., Erel, Y. & Blum, J. D. Systematic changes in lead isotopic composition with soil age in glacial granitic terrains. *Geochim. Cosmochim. Acta* **62**, 33–46 (1998).
- Erel, Y. & Morgan, J. J. The relationship between rock-derived lead and iron in natural waters. *Geochim. Cosmochim. Acta* **56**, 4157–4167 (1992).
- Harlavan, Y. & Erel, Y. The release of Pb and REE from granitoids by the dissolution of accessory phases. *Geochim. Cosmochim. Acta* **66**, 837–848 (2002).
- von Blanckenburg, F. & Nägler, T. F. Weathering versus circulation-controlled changes in radiogenic isotope tracer composition of the Labrador Sea and Northern Atlantic Deep Water. *Paleoceanography* **16**, 424–434 (2001).
- Vance, D. & Burton, K. Neodymium isotopes in planktonic foraminifera: a record of the response of continental weathering and ocean circulation rates to climate change. *Earth Planet. Sci. Lett.* **173**, 365–379 (1999).
- Reynolds, B. C., Sherlock, S. C., Kelley, S. P. & Burton, K. W. Radiogenic isotope records of Quaternary glaciations: changes in the erosional source and weathering processes. *Geology* **32**, 861–864 (2004).
- Blum, J. D. in *Tectonic Uplift and Climate Change* (ed. Ruddiman, W. F.) 259–288 (Plenum, New York, 1997).
- Millot, R., Gaillardet, J., Dupré, B. & Allègre, C. J. The global control of silicate weathering rates and the coupling with physical erosion: new insights from rivers of the Canadian Shield. *Earth Planet. Sci. Lett.* **196**, 83–98 (2002).
- Kump, L. R. & Alley, R. B. in *Material Fluxes on the Surface of the Earth* (eds Usselman, T. M. & Hay, W. W.) 46–60 (National Academy Press, Washington DC, 1994).
- Henderson, G. M., Martel, D. J., O’Nions, R. K. & Shackleton, N. J. Evolution of seawater ⁸⁷Sr/⁸⁶Sr over the last 400-Ka—the absence of glacial interglacial cycles. *Earth Planet. Sci. Lett.* **128**, 643–651 (1994).
- Catubig, N. et al. Global deep-sea burial rate of calcium carbonate during the Last Glacial Maximum. *Paleoceanography* **13**, 691–697 (1998).
- Broecker, W. S. et al. Evidence for a reduction in the carbonate ion content of the deep sea during the course of the Holocene. *Paleoceanography* **14**, 744–752 (1999).
- Reynolds, B. C., Frank, M. & O’Nions, R. K. Nd- and Pb-isotope time series from Atlantic ferromanganese crusts: implications for changes in provenance and paleocirculation over the last 8 Myr. *Earth Planet. Sci. Lett.* **173**, 381–396 (1999).
- Roy, M., Clark, P. U., Raisbeck, G. M. & Yiou, F. Geochemical constraints on the regolith hypothesis for the middle Pleistocene transition. *Earth Planet. Sci. Lett.* **227**, 281–296 (2004).
- Imbrie, J. et al. in *Milankovitch and Climate* (eds Berger, A. L., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B.) 269–305 (Reidel, Dordrecht, 1984).
- Millot, R., Allègre, C. J., Gaillardet, J. & Roy, S. Lead isotopic systematics of major river sediments: a new estimate of the Pb isotopic composition of the Upper Continental Crust. *Chem. Geol.* **203**, 75–90 (2004).
- Fagel, N. et al. Nd and Pb isotope signatures of the clay-sized fraction of Labrador Sea sediments during the Holocene: Implications for the inception of the modern deep circulation pattern. *Paleoceanography* **19**, doi:10.1029/2003PA000993 (2004).
- Farmer, L. G., Barber, D. & Andrews, J. Provenance of Late Quaternary ice-proximal sediments in the North Atlantic: Nd, Sr and Pb isotopic evidence. *Earth Planet. Sci. Lett.* **209**, 227–243 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Henderson for providing the crusts analysed in this paper; C. Hawkesworth and T. Elliott for constructive comments on an earlier draft of this manuscript; C. Coath for laboratory assistance; and B. Charlier for help with micro-drilling. Support for this research was provided by a NERC fellowship to G.L.F. and a NERC Joint Infrastructure Fund grant to the Department of Earth Sciences, University of Bristol.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.L.F. (g.l.foster@bristol.ac.uk).

LETTERS

Slip on 'weak' faults by the rotation of regional stress in the fracture damage zone

D. R. Faulkner¹, T. M. Mitchell¹, D. Healy^{1†} & M. J. Heap^{1†}

Slip on unfavourably oriented faults with respect to a remotely applied stress is well documented and implies that faults such as the San Andreas fault¹ and low-angle normal faults² are weak when compared to laboratory-measured frictional strength³. If high pore pressure within fault zones is the cause of such weakness, then stress reorientation within or close to a fault is necessary to allow sufficient fault weakening without the occurrence of hydrofracture⁴. From field observations of a major tectonic fault, and using laboratory experiments and numerical modelling, here we show that stress rotation occurs within the fractured damage zone surrounding faults. In particular, we find that stress rotation is considerable for unfavourably oriented 'weak' faults. In the 'weak' fault case, the damage-induced change in elastic properties provides the necessary stress rotation to allow high pore pressure faulting without inducing hydrofracture.

The concept of stress rotation within fault zones is important to understand the strength of fault zones that are unfavourably oriented with respect to the remotely applied driving stress ('weak' faults). The San Andreas 'stress–heat flow paradox' (where slip is implied at low shear stresses)^{1,5} and slip on low-angle normal faults^{2,6} imply that fault weakening occurs. Possible explanations include weak fault materials⁷, dynamic slip weakening^{8,9} and elevated pore fluid pressures^{4,10,11}. Stress rotation within the fault zone (as a consequence of mean stress increase, differential stress decrease and mechanical continuity) must accompany high pore fluid pressures, or effective σ_3 will be pushed well into the tensile field. This would result in hydrofracture, pore fluid pressure loss and fault strengthening⁴. Stress rotations around faults have been inferred from seismological studies^{12,13}. Previously, stress rotation has been postulated to occur in the narrow fault gouge core either by Coulomb plasticity¹⁴ or by material softening⁴. One consequence of invoking stress rotation within a deforming narrow ductile fault gouge layer is that a limiting condition occurs where σ_1 is oriented at 45° to the fault plane at the boundary of the gouge layer^{4,14}. Some workers have suggested that this soft gouge layer would be subject to extrusion as the deformation progresses¹⁵.

Faults consist of a single or multiple core zone where rock flour (gouge) accommodates the vast majority of strain, surrounded by a fractured 'damage' zone^{16,17}. We measure the density of the microfractures in the damage zone of a major strike-slip fault, and relate these to elastic property changes measured in laboratory experiments. Stress modelling shows that significant changes in the stress field occur within the fault damage zone.

We have studied the damage zone surrounding the Caleta Coloso fault, a >5-km-offset strike-slip fault in the Atacama fault system of northern Chile¹⁸. The fault cuts through crystalline rocks of predominantly granodioritic composition. The fault was exhumed and currently displays fault zone structures formed at depths of 4–10 km

that are particularly well exposed and preserved owing to the hyper-arid climate. We collected samples at regular intervals within the fracture damage zone as a function of distance from the fault core for microstructural analysis. Microfracture densities were measured (represented by fluid inclusion planes—see Fig. 1) in quartz grains within the samples. Although the microfractures within the samples are healed (fluid inclusion planes), they are likely to have remained open for significant periods of the fault's history. Hydrothermal cementation of fractures in the region shows only minor disturbance by later fault movements, supporting this inference. Open fractures in the damage zone are expected during fault movement when the fault is more or less critically stressed, as shown by well-connected fluid pathways that keep fluid pressures hydrostatic in intraplate

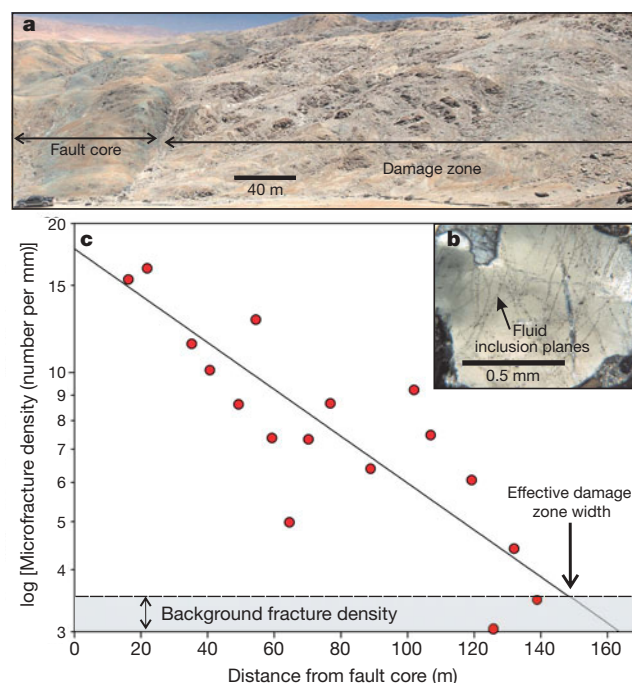


Figure 1 | Field data showing the variation in microfracture density within the fault damage zone. **a**, View looking south along the strike of the Caleta Coloso fault, showing the fault core zone to the left and fault damage zone from which samples were taken for analysis on the right. **b**, Example of quartz grain containing fluid inclusion planes. **c**, Microfracture (fluid inclusion plane) density versus distance from the fault core. The line indicates a least-squares fit to the data. Background fracture density is shown for reference.

¹Rock Deformation Laboratory, Department of Earth and Ocean Sciences, University of Liverpool, Liverpool, L69 3GP, UK. [†]Present addresses: Department of Earth Sciences, University of Durham, Durham, DH1 3LE, UK (D.H.); Mineral, Ice and Rock Physics Laboratory, Department of Earth Sciences, University College London, Gower Street, London, WC1E 6BT UK (M.J.H.).

regions¹⁹. This damage-zone permeability contrasts with the low permeability of fault cores capable of maintaining fluid overpressure¹¹.

Quartz was selected for microfracture analysis because it has little fracture anisotropy and hence is a good proxy for the total amount of damage the rock has sustained. Figure 1 shows that microfracture density falls exponentially with distance from the fault core, and is in accordance with other observations of microfracture damage surrounding lower displacement faults^{20,21}. Our microfracture densities are lower than in previous studies, perhaps owing to the smaller proportion of quartz and fracturing within feldspars. However, we suggest that the data are representative of the relative amount of microfracture damage that the rock has sustained. Having established the distribution of microfractures surrounding faults, we now consider the role of microfracturing on the elastic properties of crystalline rock from laboratory experiments.

We performed uniaxial cyclic loading, increasing stress amplitude tests on 25.4-mm-diameter cores of Westerly granite. Westerly granite was chosen for experiments in preference to the granodiorite because of fine grain size and low initial microfracture density. However, because both are low-porosity, crystalline rocks, Westerly granite should provide a representative data set that can be related to granodiorite. Details of the experiments are given in the Methods section.

Multiple tests were conducted with different numbers of increasing amplitude stress cycles so that the samples could be thin-sectioned to establish the density of microfractures as a function of number of cycles. The peak stress for the first cycle was 80 MPa and it was increased by 12 MPa in each subsequent cycle. The elastic parameters were measured in the final unloading–loading stress cycle before thin sectioning. The results are shown in Fig. 2. We chose a differential stress value of 60 MPa to observe the change in the elastic parameters resulting from damage due to the increasing number of cycles. This value is $\sim 30\%$ of the failure stress of the rock. If the rock underwent failure in nature, microfractures would form at or near the peak stress immediately before fault formation, whereupon the stress would drop to that required for frictional sliding. Microfractures may also occur as a product of irregularities on the fault, or from dynamic loading^{22,23}. Figure 2 shows a reduction in Young's modulus of ~ 6.5 GPa, and an increase in Poisson's ratio of

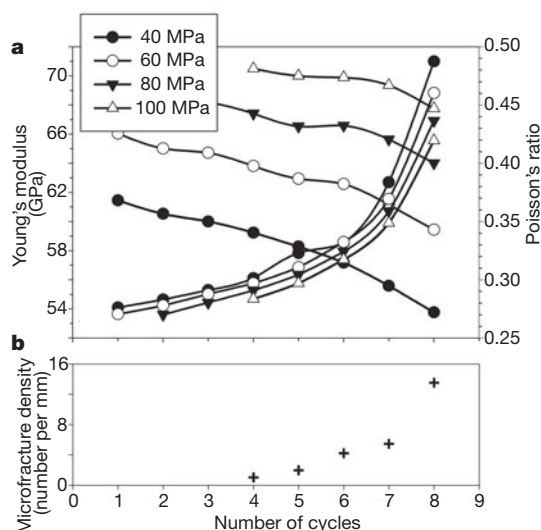


Figure 2 | Variation of measured elastic properties with increasing microfracture damage in Westerly granite. **a**, Change in Young's modulus (increasing curves) and Poisson's ratio (decreasing curves) with increasing damage. Data are shown for stresses of 40, 60, 80 and 100 MPa. Cycle 1 was taken to a maximum stress of 80 MPa and the stress was increased by ~ 12 MPa on each subsequent cycle. **b**, Increase in microfracture density with increasing damage.

~ 0.2 related to increasing microfracture density as the sample approached failure.

Despite having collected our data in the absence of confining pressure, which might serve to suppress the formation of new microfractures, we believe that the changes in the elastic parameters are representative of those at depth. Although we are not aware of any tests that have been performed to establish the static elastic parameters under confining pressure, we have compared our results to tests where dynamic elastic parameters were obtained from ultrasonic wave velocities measured during progressive fracture²⁴. The magnitudes of the change of these properties ($\Delta E \approx 10$ GPa) compare favourably to our results. Our tests were also conducted under atmospheric conditions and so we have not established the effects of water saturation and elevated fluid pressure on the results.

Thus far we have shown that microfracture density varies as a function of distance from faults and that increasing microfracture damage strongly affects the elastic properties of rocks. We now combine these data to predict the effect these changes have on a remotely applied stress field in the vicinity of a fault.

We developed a two-dimensional plane-strain model to calculate the stress field in a horizontal plane cutting a vertical strike-slip fault zone. The model differs from that of Rice⁴ in that we do not assume a rigid–plastic rheology with an actively deforming Von Mises plastic layer, but rather a statically loaded fault system with several layers of material having elastic properties that vary systematically with distance from the fault core. We justify this static case by relating it to either a locked fault in the interseismic period, or a creeping fault, where the stresses within the damage zone would be at a more or less constant level. The model is similar to that of Casey²⁵ but we represent the damage zone as a series of discrete fault-parallel layers, each with a specific Young's modulus and Poisson's ratio.

Because the laboratory experiments were performed on Westerly granite and the microfracture densities in the field were measured in

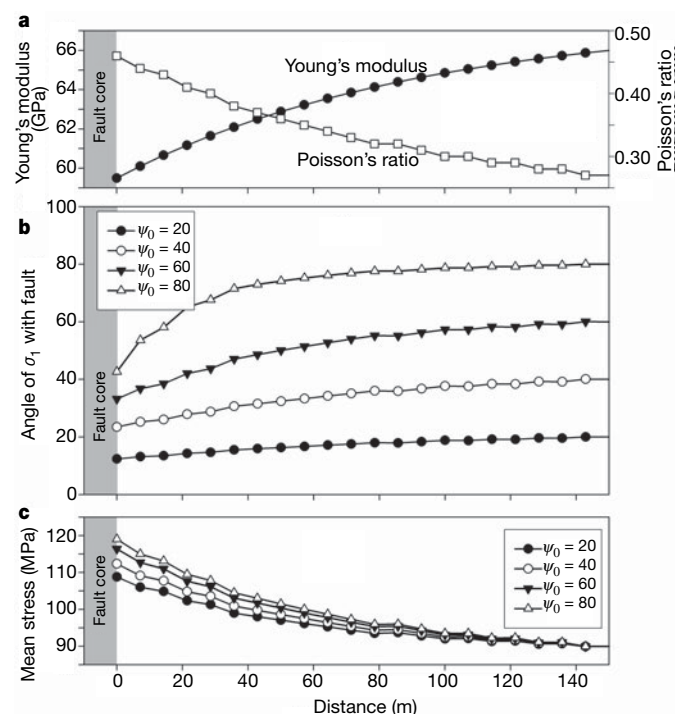


Figure 3 | Variation of elastic properties, principal stress orientation and the mean stress as a function of distance within the fault damage zone. **a**, Values for elastic properties versus distance from the fault core used for the modelling. **b**, Change in angle ψ between the fault plane and the maximum compressive stress σ_1 with distance from the fault core. ψ_0 is the value of ψ outside the damage zone. **c**, Change in the mean stress with distance from the fault core.

granodiorite, we relate the two sets of microfracture densities by assuming that the elastic properties of the Westerly granite after one stress cycle represent those of the granodiorite at the edge of the fracture damage zone, and that the elastic properties of the granite closest to failure represent those of the granodiorite adjacent to the fault core. The elastic property data are scaled according to the microfracture density. Figure 3a shows how the elastic parameters vary as a function of distance from the fault core. In cartesian coordinates, with x parallel to the fault and y normal to the fault, we imposed the following boundary conditions for stress σ and strain ε for cohesive (no slip) contacts between each layer n :

$$(\sigma_{yy})_n = (\sigma_{yy})_{n+1}$$

$$(\tau_{yx})_n = (\tau_{yx})_{n+1}$$

$$(\varepsilon_{xx})_n = (\varepsilon_{xx})_{n+1}$$

Assuming mechanical equilibrium, the equations relating normal and shear stress and strain for each layer could then be solved in the same way as in ref. 25. The calculated cartesian stress components were then converted into principal stresses using the appropriate transformation equations²⁶.

We selected remotely applied principal stress values of 120 and 60 MPa. Using these values gives a differential stress of 60 MPa: the same differential stress at which we measured our elastic parameters in the laboratory. The model results are shown in Fig. 3b and c. A remotely applied σ_1 oriented at 80° to the fault plane will rotate in the damage zone such that the fault core will only 'see' local stresses that are oriented at an angle of 42° to the fault plane. The amount of rotation is less for smaller angles between σ_1 and the fault plane. The mean stress, as well as the values of σ_1 and σ_3 , increases as the fault core is approached, with an overall decrease in the differential stress (Figs 3c and 4). The Poisson's ratio has a larger effect on the rotation of stress than the Young's modulus. These trends have been predicted previously⁴ for stress rotations within a 'weak' gouge layer but only in a qualitative way. In our model, the stress imposed on the fault core has already been rotated throughout the damage zone, and we have quantified the magnitude of the stress rotation and modification of the principal stresses using field and laboratory data.

The results from our model are applicable to all brittle faults formed within low-porosity rocks where microfracture damage

and dilatancy accumulate during deformation. If a fault zone has an associated fracture damage zone, then stresses will be modified and rotated according to the change in elastic properties of the damaged material. Faults running through lithologies different from that described here might undergo different quantities of stress modification, but the trend should be equivalent. In Fig. 3, we have placed a scale on the abscissa to relate our data specifically to its distance from the Caleta Coloso fault core. However, our model is scale invariant and so lower displacement faults, perhaps with smaller damage zones, should still undergo similar stress modification if the elastic property changes are commensurate with those we measured. The type of fault zone, either with a single¹⁶ or multiple core¹⁷ should not affect the surrounding damage zone. With regard to fault gouge extrusion¹⁵, we note that the fault core in large faults is only a small proportion of the width of the fault zone^{16,17}. Hence only minor extrusion would be possible before asperities would contact, and because our model does not rely on stress modification in the gouge zone, stress changes within the damage zone would still allow pore pressure weakening without hydrofracture.

In this paper we consider only the effect of microfractures on the bulk elastic properties of the rock. In fault damage zones, macroscopic fractures are present but their effect is on too large a scale for laboratory measurements and hence are not discussed here. However, the presence of macrofractures will further affect the bulk elastic properties and hence our analysis could be thought of as the minimum effect that might be seen.

We emphasize that although stresses are rotated to favourable angles, stress rotation alone does not bring an unfavourably oriented fault closer to failure^{4,25}. One of our model boundary conditions is that the normal and shear stresses between each fault-parallel layer are equal. Figure 4 shows the modification of the stress field in the fracture damage zone, and the fixed normal and shear stresses resolved parallel to the fault plane. For failure to occur, fault zone weakening is still required, but stress rotation within the damage zone allows high pore fluid pressure to develop without hydrofracture and rupture of any hydraulic seal present. Our model is quantitatively based on field and laboratory measurements and circumvents the problems associated with invoking stress rotation within the narrow gouge layer of the fault core.

METHODS

In the experiments, the axial and circumferential strain were both measured as a function of stress during loading and unloading and from these data the elastic parameters, Young's modulus E and Poisson's ratio ν , were calculated. The stress-strain curves look like typical uniaxial stress-strain curves, but with multiple cycles. Unlike previous uniaxial compression tests that have been used to determine the static elastic properties of rocks, we have extended the range of stress values for which values of the elastic properties can be obtained. In the post-yield phase of loading, the loading-unloading curves display hysteresis. We worked around this by fitting third-order polynomial functions (to capture the true form of the stress-strain curve) to the loading and unloading stress strain curves that can then be differentiated to obtain the gradient. The gradients of an unloading curve were then compared with the gradients for the next loading curve (to ensure the rock did not incur any further microfracture damage between the two curves until the previous highest stress in the unloading curve was exceeded), and the range of stress values for which the gradients were the same were then used to calculate the elastic parameters²⁷. In this way, elastic constants were obtained over a wide range of stress values.

Our data show that the Poisson's ratio does not extend above 0.5—an indication of the validity of our approach. Our approach means that the elastic parameters immediately before failure are not recorded and are likely to have a greater range than those measured, and hence the data we use for modelling may be seen as conservative, lower-bound values for the elastic properties. Thus the effects described later are a minimum response of the stress orientation and magnitude.

Received 25 May; accepted 12 October 2006.

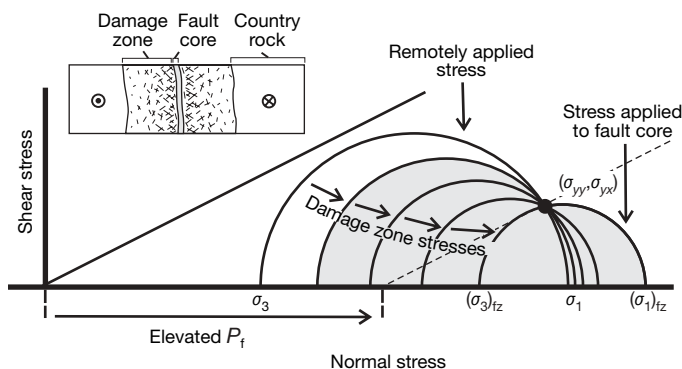


Figure 4 | Mohr diagram illustrating the stress rotation within the damage zone. Schematic Mohr diagram showing the change in the state of stress within the damage zone (shaded circles) to the fault core (fz). The two failure envelopes show the case of an unpressurized fault core (solid line) and pressurized fault core (dashed line), where P_f denotes the pore fluid pressure. Note that the normal stress and shear stress resolved on the fault plane are fixed and do not change from the remotely applied stress to the fault plane. The inset shows a schematic diagram of fault zone structure illustrating the core zone, damage zone and country rock.

1. Zoback, M. D. *et al.* New evidence on the state of stress of the San Andreas fault system. *Science* **238**, 1105–1111 (1987).

2. Colletini, C. & Sibson, R. H. Normal faults, normal friction? *Geology* **29**, 927–930 (2001).
3. Byerlee, J. D. Friction of rocks. *Pure Appl. Geophys.* **116**, 615–629 (1978).
4. Rice, J. R. in *Fault Mechanics and Transport Properties of Rocks* (eds Evans, B. & Wong, T.-f.) 475–503 (Academic Press, San Diego, CA, 1992).
5. Lachenbruch, A. H. & Sass, J. H. Heat flow and energetics of the San Andreas fault zone. *J. Geophys. Res.* **85**, 6185–6223 (1980).
6. Sibson, R. H. An assessment of field evidence for “Byerlee” friction. *Pure Appl. Geophys.* **142**, 645–662 (1993).
7. Morrow, C. A., Moore, D. E. & Lockner, D. A. The effect of mineral bond strength and absorbed water on fault gouge frictional strength. *Geophys. Res. Lett.* **27**, 815–819 (2000).
8. Wibberley, C. A. J. & Shimamoto, T. Earthquake slip weakening and asperities explained by thermal pressurization. *Nature* **436**, 689–692 (2005).
9. Di Toro, G., Hirose, T., Nielsen, S., Pennacchioni, G. & Shimamoto, T. Natural and experimental evidence of melt lubrication of faults during earthquakes. *Science* **311**, 647–650 (2006).
10. Byerlee, J. D. Friction, overpressure and fault-normal compression. *Geophys. Res. Lett.* **17**, 2109–2112 (1990).
11. Faulkner, D. R. & Rutter, E. H. Can the maintenance of overpressured fluids in large strike-slip fault zones explain their apparent weakness? *Geology* **29**, 503–506 (2001).
12. Hardebeck, J. L. & Hauksson, E. Role of fluids in faulting inferred from stress field signatures. *Science* **285**, 236–239 (1999).
13. Provost, A.-S. & Houston, H. Orientation of the stress field surrounding the creeping section of the San Andreas fault: Evidence for a narrow mechanically weak fault zone. *J. Geophys. Res.* **106**, 11373–11386 (2001).
14. Lockner, D. A. & Byerlee, J. D. How geometrical constraints contribute to the weakness of mature faults. *Nature* **363**, 250–252 (1993).
15. Scholz, C. H. Evidence for a strong San Andreas fault. *Geology* **28**, 163–166 (2000).
16. Chester, F. M., Evans, J. P. & Biegel, R. L. Internal structure and weakening mechanisms of the San Andreas fault. *J. Geophys. Res.* **98**, 771–786 (1993).
17. Faulkner, D. R., Lewis, A. C. & Rutter, E. H. On the internal structure and mechanics of large strike-slip faults: field observations from the Carboneras fault, southeastern Spain. *Tectonophysics* **367**, 235–251 (2003).
18. Cembrano, J. et al. Fault zone development and strain partitioning in an extensional strike-slip duplex: a case study from the Atacama Fault System, northern Chile. *Tectonophysics* **400**, 105–125 (2005).
19. Townend, J. & Zoback, M. D. How faulting keeps the crust strong. *Geology* **28**, 399–402 (2000).
20. Vermilye, J. & Scholz, C. H. The process zone; a microstructural view of fault growth. *J. Geophys. Res.* **103**, 12223–12237 (1998).
21. Anders, M. H. & Wiltschko, D. V. Microfracturing, paleostress and the growth of faults. *J. Struct. Geol.* **16**, 795–815 (1994).
22. Wilson, J. E., Chester, J. S. & Chester, F. M. Microfracture analysis of fault growth and wear processes, Punchbowl Fault, San Andreas system, California. *J. Struct. Geol.* **25**, 1855–1873 (2003).
23. Rice, J. R., Sammis, C. G. & Parsons, R. Off-fault secondary failure induced by a dynamic slip pulse. *Bull. Seismol. Soc. Am.* **95**, 109–134 (2005).
24. Ayling, M. R., Meredith, P. G. & Murrell, S. A. F. Microcracking during triaxial deformation of porous rocks monitored by changes in rock physical properties. I. Elastic-wave propagation measurements on dry rocks. *Tectonophysics* **245**, 205–221 (1995).
25. Casey, M. Mechanics of shear zones in isotropic dilatant materials. *J. Struct. Geol.* **2**, 143–147 (1980).
26. Jaeger, J. C. & Cook, N. G. W. *Fundamentals of Rock Mechanics* 11–17 (Chapman and Hall, London, 1976).
27. Heap, M. J. & Faulkner, D. R. Quantifying the evolution of static elastic properties as crystalline rock approaches failure. *J. Struct. Geol.* (submitted)

Acknowledgements We thank J. Cembrano for field assistance and M. Casey, N. Kusznir and D. Prior for discussions. This work was supported by the Natural Environment Research Council and the Nuffield Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.R.F. (faulkner@liv.ac.uk).

LETTERS

Dynamical evolution of ecosystems

Sandro Azae¹, Simone Pigolotti², Jayanth R. Banavar³ & Amos Maritan¹

The assembly of an ecosystem such as a tropical forest depends crucially on the species interaction network, and the deduction of its rules is a formidably complex problem¹. In spite of this, many recent studies^{2–16} using Hubbell's neutral theory of biodiversity and biogeography² have demonstrated that the resulting emergent macroscopic behaviour of the ecosystem at or near a stationary state shows a surprising simplicity reminiscent of many physical systems¹⁷. Indeed the symmetry postulate², that the effective birth and death rates are species-independent within a single trophic level, allows one to make analytical predictions for various static distributions such as the relative species abundance^{3–12}, β -diversity^{13–15} and the species–area relationship¹⁶. In contrast, there have only been a few studies of the dynamics and stability of tropical rain forests^{18–20}. Here we consider the dynamical behaviour of a community, and benchmark it against the exact predictions of a neutral model near or at stationarity. In addition to providing a description of the relative species abundance, our analysis leads to a quantitative understanding of the species turnover distribution and extinction times, and a measure of the temporal scales of neutral evolution. Our model gives a very good description of the large quantity of data collected in Barro Colorado Island in Panama in the period 1990–2000 with just three ecologically relevant parameters and predicts the dynamics of extinction of the existing species.

We present an analytical model that allows one to probe the characteristic timescales of evolving tropical forests and to evaluate the consequences of anthropogenic processes. Our approach is valid for an ecosystem at or near stationarity; indeed, one would expect important deviations from our predictions when the stationarity assumption is not valid (see, for example, ref. 21). Using a neutral model, we have obtained exact solutions for the probability distribution, $P(x, t)$, that a species has a population x at time t for arbitrary initial and boundary conditions (see Supplementary Information for details). The species are assumed to be non-interacting and are characterized by effective birth and death rates given by $b(x) = b_1x + b_0$ and $d(x) = d_1x + d_0$ respectively, where b_1 and d_1 are the per-capita rates and the constants b_0 and d_0 incorporate density dependence and result in a rare species advantage when $b_0 > d_0$ (ref. 9). To simplify the analytical treatment and for parsimony we have chosen $b_0 = -d_0$ in our analysis.

There are three biological parameters in our framework, namely τ , b and D : τ is the characteristic timescale associated with species turnover in neutral evolution—an ecosystem close to the stationary state is able to recover from a perturbation on a timescale of order τ and its inverse is simply the difference between d_1 and b_1 ; $b = 2b_0$ takes into account density dependence effects⁹ arising from immigration²² and/or speciation, for example; and D accounts for demographic stochasticity and is given by $(b_1 + d_1)/2$.

The steady-state solution, which is independent of initial conditions, provides an exact expression for the relative species abundance (RSA),

$$P_{\text{RSA}}(x) = \frac{(D\tau)^{-b/D}}{\Gamma(b/D)} x^{b/D-1} e^{-x/D\tau} \quad (1)$$

($\Gamma(x)$ is the gamma function²³), which is in good accord with RSA data for various censuses of the Barro Colorado Island (BCI) forest (Center for Tropical Forest Science website, <http://ctfs.si.edu>), and for several other tropical forests with the use of the data presented in the Supplementary Information of ref. 9. These fits allow us to estimate two combinations, b/D and $D\tau$ (see Fig. 1 and Supplementary Tables 1 and 2), of the three parameters.

The time-dependent species turnover distribution (STD), defined as the probability $P_{\text{STD}}(\lambda, t)$ that the ratio of the populations of a species separated by a time interval t , $x(t)/x(0)$, is equal to λ , is found under stationary conditions to be

$$P_{\text{STD}}(\lambda, t) = A \frac{(\lambda + 1)}{\lambda} \frac{(e^{t/\tau})^{b/2D}}{1 - e^{-t/\tau}} \left(\frac{\sinh(t/2\tau)}{\lambda} \right)^{b/D+1} \left(\frac{4\lambda^2}{(\lambda + 1)^2 e^{t/\tau} - 4\lambda} \right)^{b/D+1/2} \quad (2)$$

where A is the normalization constant. P_{STD} depends only on b/D and τ . The above result, derived for reflecting boundary conditions at

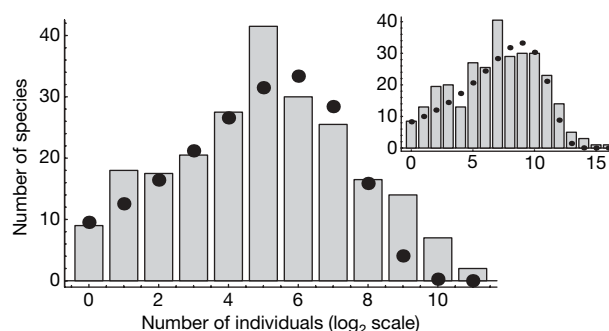


Figure 1 | Relative species abundance plot in the BCI forest from the 1990 census (Center for Tropical Forest Science website). The individuals of more than 10 cm d.b.h. in this tropical forest are binned with the method of refs 7, 29. The inset shows the same histogram for the individuals of more than 1 cm d.b.h. for the same forest and yields consistent estimates of the model parameters and temporal scales within the error bars. The estimated parameters are robust within error bars on changing the nature of binning of the empirical data to non-overlapping bins. The points are the best fits to the mean number of species with population between 2^{n-1} and 2^n , as given by equation (1). The fit for large x is readily improved at the cost of introducing an additional parameter (see Supplementary Information for error analysis and other details). Note that the RSA plot for individuals of more than 10 cm d.b.h. is smoother at low abundance than the plot for individuals of more than 1 cm d.b.h. This is to be expected because younger populations are subject to larger fluctuations than older ones.

¹Dipartimento di Fisica 'G. Galilei', Università di Padova, via Marzolo 8, 35131 Padova, Italy. ²The Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark. ³Department of Physics, The Pennsylvania State University, 104 Davey Laboratory, University Park, Pennsylvania 16802, USA.

$x = 0$, is essentially independent of the specific boundary conditions for the relatively short time interval of 5–10 years for which the dynamical data exists. The histograms in Fig. 2 are fitted with equation (2), with τ as a free parameter (see Supplementary Tables 1 and 2). We find τ , the characteristic timescale for the BCI forest, to be about $3,500 \pm 1,000$ years (for trees of more than 10 cm stem diameter at breast height (d.b.h.)) and $2,900 \pm 1,100$ years (for trees of more than 1 cm d.b.h.), where the broad uncertainty is due to the fact that the data are sampled at relatively short time intervals. Long-term vegetation studies of a *Tsuga*-dominated forest in New England²⁴ reveal a recovery time for the *Tsuga canadensis* that is of the same order of magnitude (roughly 1,500 years) in spite of being quite different from the BCI forest. Studies of Pleistocene forest dynamics²⁵ are also in good agreement with our estimate. The exact analytic expression permits the prediction of a characteristic timescale more than two orders of magnitude larger than the time interval of measurement, reminiscent of the classic example of the prediction of the half-lives of radioactive isotopes, which are often many times the age of the Universe. The estimate of the timescale becomes asymptotically exact as the number of species or nuclei increases. The BCI data analysis for trees of more than 10 cm d.b.h. yields $b = 0.02 \pm 0.01$ and $D = 0.04 \pm 0.02$ (both in units of individuals per year), leading to an effective per-capita death rate $d_1 = (1/\tau + 2D)/2 \approx D$, which is consistent with the estimate given in ref. 19.

One can obtain a measure of the mean lifetime of a species, $\langle t \rangle = -\tau[\gamma + \psi(1 - b/D)]$, where the Euler constant $\gamma = 0.577\dots$, $\psi(z) = \Gamma'(z)/\Gamma(z)$ and $\Gamma(z)$ is the gamma function²³, yielding $\langle t \rangle = 3,400 \pm 1,000$ years, of the same order as τ . This timescale is in accord with the estimates of extinction times given in ref. 26. $\langle t \rangle/\tau$ depends only on b/D , which can be calculated from the steady-state RSA without the need for dynamical data. We have used equation (1) to fit the RSA of various tropical forests, yielding $\langle t \rangle/\tau = 1.94, 1.67, 0.67, 0.95$ and 1.38 for Yasuni, Lambir, Sinharaja, Korup and Pasoh, respectively (see Supplementary Information). The similarity of the two timescales $\langle t \rangle$ and τ is not mandated by the theory because the function $\psi(1 - b/D)$ diverges as b/D approaches 1. Were $\tau \gg \langle t \rangle$, one would obtain a contradiction because the ecosystem would not reach a steady state: extinction would be much faster than recovery. In contrast, $\tau \ll \langle t \rangle$ would lead to a very robust and essentially unchanging ecosystem, rapidly recovering (with respect to the extinction time) from disturbances and with little room for evolution. Our result suggests that ecosystems at stationarity are marginally

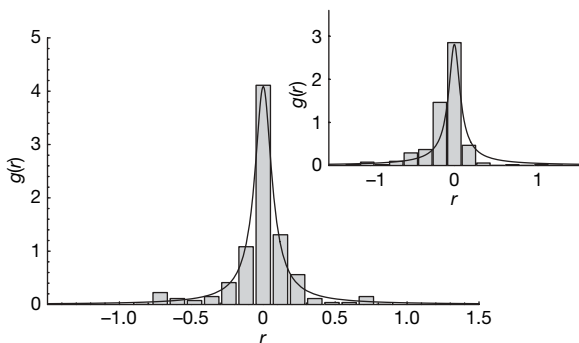


Figure 2 | STD for the interval 1990–95 in the BCI forest. The main panel shows results for individuals of more than 10 cm d.b.h., and the inset results for individuals of more than 1 cm d.b.h. (Center for Tropical Forest Science website). We have defined the new variable $r = \log(\lambda)$, which is distributed as $g(r) = e^r P_{STD}(e^r, t)$, where $P_{STD}(\lambda, t)$ is given by equation (2). Data are plotted with a linear binning in the $r = \log(\lambda)$ axis and fitted to $g(r)$. b/D is obtained from fitting the RSA data in 1990 (see Fig. 1). The best-fit parameter is found to be $\tau \sim 4,400$ years for individuals of more than 10 cm d.b.h., and $\tau \sim 3,900$ years for individuals of more than 1 cm d.b.h. The fits of both RSA (see Fig. 1) and STD for individuals of more than 10 cm d.b.h. are systematically better than those for individuals of more than 1 cm d.b.h.

stable—not so stable that they are frozen in time and not so fragile that they are prone to extinction.

One may also study the time correlation function $\kappa(t) = \langle x(t)x(0) \rangle - \langle x(t) \rangle \langle x(0) \rangle$, where the averages are taken over the species. For neutral dynamics, we find that $\kappa(t) = bD\tau^2 e^{-t/\tau}$. Physically, the time correlation function approaches zero in an exponential fashion, with τ as the characteristic relaxation time. Although the currently available data for the BCI forest are not sufficient for analysis, a study of the time correlation function provides a simple yet powerful probe for obtaining the timescale τ as well as for probing departures from neutrality.

We turn now to a prediction of a quantity that we call the restricted relative species abundance, $q(x, t)$, which is a measure of the relative species abundance of just the species present in the forest at initial time denoted by $t = 0$. Unlike the RSA, the restricted relative species abundance does not consider any new species or any species reintroduced after its original extinction. One finds that

$$q(x, t) = \frac{\alpha^\beta x^{\beta-1} e^{-\alpha x}}{\Gamma(\beta) \Gamma(1-\beta)} \gamma\left(1-\beta, \frac{\alpha x}{e^{t/\tau}-1}\right) \quad (3)$$

where $\Gamma(x)$ is the gamma function and $\gamma(a, z) = \int_0^z t^{a-1} e^{-t} dt$ (see ref. 23), $\beta = b/D$, and $\alpha = 1/D\tau$. Figure 3 depicts the inexorable march to extinction of the species, with the rare species being more prone to extinction. Note that, unlike the RSA and STD, the fitting of restricted relative species abundance, when the data become available, can be employed to determine all three parameters of the model simultaneously. Unfortunately, with just one set of data (RSA or STD), it is not possible to determine the parameters and predict the behaviour of the other in a manner similar to the treatment in ref. 27.

We conclude by noting that different choices of b_0 and d_0 have also been shown to lead to very good fits of the RSA of various tropical forests, but the analytical treatment for the dynamics is much more involved. Indeed, the predicted macroscopic properties of the ecosystem at or near stationarity are not very sensitive to specific choices of the model parameters but rather depend on the key assumptions; that is, the symmetry (or neutrality) principle and the independent individual effective dynamics. Our approach represents a complete description of an ecosystem undergoing neutral dynamics that can be solved analytically. The goodness of the fits of both RSA and STD does not imply that neutrality is a valid description of how nature operates. Rather, the analytical expressions for various measurable

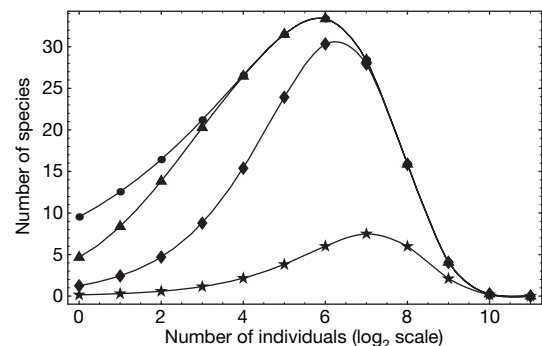


Figure 3 | Restricted relative species abundance. Plot of the mean number of species originally present in an ecosystem with population between 2^{n-1} and 2^n after time t has evolved, as given by equation (3). The circles denote the steady state at $t = 0$, namely the standard RSA; the triangles correspond to $t = 100$ years; the diamonds to $t = 1,000$ years; and the stars to $t = 10,000$ years. The parameters are those deduced from the RSA of the BCI plot in 1990 for more than 10 cm d.b.h. (see Supplementary Table 1). Note the shift of the maximum of the curve to the right and that rare species are more prone to extinction.

quantities will allow one not only to estimate the relevant timescales but also to look for deviations from the idealized theory. It yields good fits of both RSA and STD, it provides estimates of timescales that are in accord with previous analysis and it also predicts the extinction-time distribution. The issue of model selection²⁸ is beyond the scope of our current analysis and will become increasingly important as other models with exact solutions for the dynamics are developed. Our model is an ideal starting point for the incorporation of features such as non-neutral dynamics and time-dependent environmental and spatial effects including heterogeneities.

Received 14 August; accepted 5 October 2006.

- Montoya, J. M., Pimm, S. L. & Solé, R. Ecological networks and their fragility. *Nature* **442**, 259–264 (2006).
- Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ., New Jersey, 2001).
- Bell, G. Neutral macroecology. *Science* **293**, 2413–2418 (2001).
- Chave, J. Neutral theory and community ecology. *Ecol. Lett.* **7**, 241–253 (2004).
- Bell, G. The distribution of abundance in neutral communities. *Am. Nat.* **155**, 606–617 (2000).
- McKane, A., Alonso, D. & Solé, R. V. Mean-field stochastic theory for species-rich assembled communities. *Phys. Rev. E* **62**, 8466–8484 (2000).
- Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. Neutral theory and relative species abundance in ecology. *Nature* **424**, 1035–1037 (2003).
- Pigoletti, S., Flammini, A. & Maritan, A. Stochastic model for the species abundance problem in an ecological community. *Phys. Rev. E* **70**, 011916 (2004).
- Volkov, I., Banavar, J. R., He, F., Hubbell, S. P. & Maritan, A. Density dependence explains tree species abundance and diversity in tropical forests. *Nature* **438**, 658–661 (2005).
- Alonso, D. & McKane, A. Sampling Hubbell's neutral theory of biodiversity. *Ecol. Lett.* **7**, 911–914 (2004).
- Alonso, D., Etienne, R. S. & McKane, A. J. The merits of neutral theory. *Trends Ecol. Evol.* **21**, 451–457 (2006).
- Etienne, R. S. & Alonso, D. A dispersal-limited sampling theory for species and alleles. *Ecol. Lett.* **8**, 1147–1156 (2005).
- Chave, J. & Leigh, E. G. Jr. A spatially-explicit model of β -diversity. *Theor. Popul. Biol.* **62**, 153–168 (2002).
- Condit, R. *et al.* Beta-diversity in tropical forest trees. *Science* **295**, 666–669 (2002).
- Zillio, T., Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. Spatial scaling in model plant communities. *Phys. Rev. Lett.* **95**, 098101 (2005).
- Durrett, R. & Levin, S. A. Spatial models for species area curves. *J. Theor. Biol.* **179**, 119–127 (2002).
- Harte, J. Tail of death and resurrection. *Nature* **424**, 1006–1007 (2003).
- Clark, J. S. & McLachlan, J. S. Stability of forest biodiversity. *Nature* **423**, 635–638 (2003).
- Condit, R. *et al.* Dynamics of the forest communities at Pasoh and Barro Colorado: comparing two 50-ha plots. *Phil. Trans. R. Soc. Lond. B* **354**, 1739–1748 (1999).
- Sheil, D., Jennings, S. & Savill, P. Long-term plot observations of vegetation dynamics in Budongo, a Uganda rain forest. *J. Trop. Ecol.* **16**, 765–800 (2000).
- Gilbert, B., Laurance, W. F., Leigh, E. G. Jr & Nascimento, H. E. M. Can neutral theory predict the responses of amazonian tree communities to forest fragmentation? *Am. Nat.* **168**, 304–317 (2006).
- Gonzalez, A., Lawton, J. H., Gilbert, F. S., Blackburn, T. M. & Evans-Freke, I. I. Metapopulation dynamics, abundance, and distribution in a microsystem. *Science* **281**, 2045–2047 (1998).
- Abramowitz, M. & Stegun, I. A. *Handbook of Mathematical Functions* (National Bureau of Standards, Gaithersburg, Maryland, 1964).
- Foster, D. R. & Zebrick, T. M. Long-term vegetation dynamics and disturbance history of a *Tsuga*-dominated forest in New England. *Ecology* **74**, 982–998 (1993).
- Bush, M. B., Silman, M. R. & Urrego, D. H. 48,000 years of climate and forest change in a biodiversity hot spot. *Science* **303**, 827–829 (2004).
- Pimm, S. L., Russell, G. J., Gittleman, J. L. & Brooks, T. M. The future of biodiversity. *Science* **269**, 347–350 (1995).
- Adler, P. B. Neutral models fail to reproduce observed species–area and species–time relationships in Kansas grasslands. *Ecology* **85**, 1265–1272 (2004).
- Hilborn, R. & Mangel, M. *The Ecological Detective. Confronting Models with Data* (Princeton Univ. Press, Princeton, New Jersey, 1997).
- Preston, F. W. The commonness and rarity of species. *Ecology* **29**, 254–283 (1948).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank I. Volkov for comments. This work was supported by COFIN 2005, the NSF, NASA and NSF IGERT.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.R.B. (banavar@psu.edu) or A.M. (maritan@pd.infn.it).

Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein

Shimon Bershtein¹, Michal Segal¹, Roy Bekerman¹, Nobuhiko Tokuriki¹ & Dan S. Tawfik¹

The distribution of fitness effects of protein mutations is still unknown^{1,2}. Of particular interest is whether accumulating deleterious mutations interact, and how the resulting epistatic effects shape the protein's fitness landscape. Here we apply a model system in which bacterial fitness correlates with the enzymatic activity of *TEM-1* β -lactamase (antibiotic degradation). Subjecting *TEM-1* to random mutational drift and purifying selection (to purge deleterious mutations) produced changes in its fitness landscape indicative of negative epistasis; that is, the combined deleterious effects of mutations were, on average, larger than expected from the multiplication of their individual effects. As observed in computational systems^{3–5}, negative epistasis was tightly associated with higher tolerance to mutations (robustness). Thus, under a low selection pressure, a large fraction of mutations was initially tolerated (high robustness), but as mutations accumulated, their fitness toll increased, resulting in the observed negative epistasis. These findings, supported by FoldX stability computations of the mutational effects⁶, prompt a new model in which the mutational robustness (or neutrality) observed in proteins, and other biological systems, is due primarily to a stability margin, or threshold, that buffers the deleterious physico-chemical effects of mutations on fitness. Threshold robustness is inherently epistatic—once the stability threshold is exhausted, the deleterious effects of mutations become fully pronounced, thereby making proteins far less robust than generally assumed.

Interactions between mutations have been studied by a number of fields, although nomenclature differs; geneticists term interactions within, or between genes, as 'epistasis'⁷, whereas protein biophysicists use 'double mutant cycles'⁸ to describe intragenic interactions. Because the ultimate impact of mutations depends both on the effect of each mutation on fitness and on the interactions between accumulating mutations, insights from these different disciplines can nevertheless be combined in one model².

Taking the converse view, if deleterious mutations do not interact, then under no selection, fitness (or the stability, or activity, of a protein) should decline exponentially:

$$W_{(n)} = \exp(-\alpha n) \quad (1)$$

where n is the number of mutations and α is the exponential decline parameter^{3,9}. If, however, deleterious mutations interact so that their combined impact on fitness is greater than expected from multiplying their individual effects (or more than additive in terms of $\log W$, or $\Delta\Delta G$ for protein stability and function), fitness decline would accelerate with the accumulation of mutations, giving rise to 'negative epistasis':

$$W_{(n)} = \exp(-\alpha n - \beta n^2) \quad (2)$$

where α reflects the fraction of multiplicative deleterious mutations and β is the epistasis parameter⁹.

Because neither of these crucial factors (α , β) is quantitatively understood—particularly at the level of a single protein^{1,2}—we set up an experimental system that measured the fitness landscape of a protein subjected to a lengthy random drift (up to 20 mutations per gene). We derived the exponential decline and epistasis parameters and examined their impact on the rate and dynamics of mutation accumulation.

As our model protein we chose *TEM-1* β -lactamase, an enzyme that degrades the antibiotic ampicillin and thereby confers ampicillin resistance on Gram-negative bacteria such as *Escherichia coli*. The fitness of a *TEM-1* variant is directly correlated with the maximal concentration of ampicillin that *E. coli* carrying this gene variant can tolerate. At the level of a population, fitness (W) refers to the fraction of variants that confer resistance at a given concentration of ampicillin. By determining the fraction of viable variants over the entire range of ampicillin concentrations, a fitness landscape was obtained (Supplementary Fig. 1).

The *TEM-1* gene was cloned into a plasmid (as it occurs in nature) under its endogenous promoter. Recloning after each round of mutagenesis confined the mutational drift to the open reading frame of *TEM-1*. Our *in vitro* random mutagenesis protocol was optimized for high reproducibility and was calibrated to obtain, on average, two mutations per gene per round of mutagenesis. We maintained three populations of randomly drifting *TEM-1* genes: one population under no selection (*Lib0*), and the rest under purifying selection at 'high' and 'low' stringencies. Each population, or plasmid library, was separately mutated, ligated into an empty vector and transformed into *E. coli* host cells; it then underwent purifying selection: 'high' selection pressure (250 $\mu\text{g ml}^{-1}$ ampicillin; *Lib250*; Supplementary Fig. 2), and 'low' selection pressure (12.5 $\mu\text{g ml}^{-1}$ ampicillin; *Lib12.5*). After growth on selection plates, plasmid DNA was extracted from the surviving *E. coli* colonies, and the *TEM-1* genes were subjected to the next round of mutagenesis. Altogether, ten successive rounds of mutagenesis and purifying selection were performed. Loss of diversity was less than 50% per round, and a diversity of at least 10^6 variants per library was maintained throughout.

As expected, a rapid fitness decline was observed in *Lib0* (no selection). The fitness of the selected populations (*Lib12.5* and *Lib250*) remained unchanged under the threshold of selection, and decreased above that threshold (Supplementary Fig. 3). Fitness values (W) of the randomly drifting population (*Lib0*) were plotted as a function of the mutational load n . These plots showed consistent and significant deviations from simple exponential decays. This was particularly true at low fitness levels; indeed, at less than 1,000 $\mu\text{g ml}^{-1}$ ampicillin the data could only be fitted reliably to equation (2) (Fig. 1). The 'net' effect of deleterious mutations accumulating randomly in the

¹Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel.

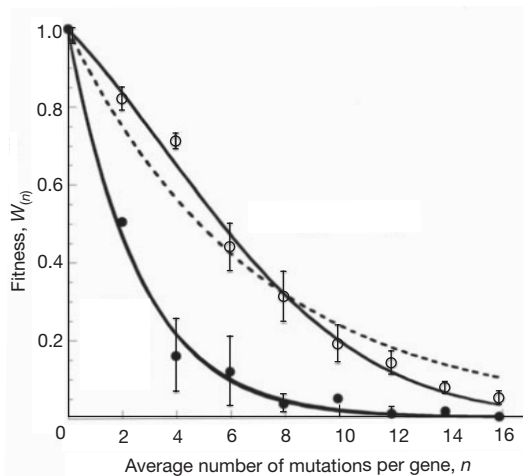


Figure 1 | Negative epistasis underlines the random drift of *TEM-1*. The fitness decline of unselected library *Lib0* at the highest fitness level (2,500 µg ml⁻¹ ampicillin; filled circles) is exponential and fits equation (1) with $\alpha = 0.371$ (lower black line). At the lowest ampicillin concentration (12.5 µg ml⁻¹; open circles) the data fit equation (2) well with $\alpha = 0.072$ and $\beta = 0.009$ (upper black line), but poorly to equation (1) ($\alpha = 0.145$; broken line). Error bars represent s.d. for two to five independent measurements of fitness.

TEM-1 gene was therefore found to be synergistic ($\beta > 0$; negative epistasis).

The α and β parameters deduced from fitness decay plots for all ampicillin concentrations (Supplementary Fig. 4) are plotted in Fig. 2a. The exponential decay parameter of *TEM-1* (α) increased almost linearly with fitness level. Extrapolating α to zero ampicillin indicated that the fraction of *TEM-1* mutations that are unconditionally lethal is about 7% (Fig. 2a), corresponding to mutations that severely undermine the stability of *TEM-1* or abolish its catalytic activity. The degree of negative epistasis was tightly correlated with fitness levels; a higher decline in fitness (larger α values) gave rise to weaker epistasis (β/α) (Fig. 2b). It therefore seems that when high fitness levels are maintained, the fraction of lethal mutations is also high (large α , low robustness). However, the remaining fraction of mutations is largely neutral, and their epistatic potential is minimal ($\beta/\alpha \approx 0$). Conversely, under low fitness levels a much larger fraction

of mutations is tolerated (small α , high robustness), but these exhibit large negative epistatic effects ($\beta/\alpha \gg 0$).

In terms of protein stability and function ($\Delta\Delta G$), the negative epistasis observed here implies that *TEM-1* mutations, on average, interact in a more-than-additive manner. However, this result does contradict our current knowledge of proteins. 'Double mutant cycles'⁶ is a commonly used tool to dissect proteins, where the effects of different mutations on various physico-chemical properties (thermodynamic stability, binding affinity or enzymatic rates) are measured on their own and together⁸. They therefore comprise a measure of epistasis. Numerous cycles performed on many different proteins indicate that the physico-chemical effects of deleterious mutations (in terms of $\Delta\Delta G$, or $\log W$) are mostly additive, or less than additive for interacting residues. Effects that are more than additive are rare exceptions⁸. We therefore surmise that the negative epistasis observed here is the outcome of a non-additive decline in fitness, in response to additive physico-chemical changes (see also ref. 2).

To examine this hypothesis we assumed that most deleterious mutations had undermined stability, thereby reducing the levels of soluble, active protein^{2,10}. We then applied the FoldX⁶ algorithm to predict the stability changes induced by mutations in the drifting *TEM-1* populations. FoldX analysis indicated that the destabilizing mutations frequently observed in the unselected library *Lib0* were purged by the purifying selections (Fig. 3a). Initially (as indicated by sampling of the fifth round), mutations with considerable destabilizing effects ($\Delta\Delta G \leq 6.5$ kcal mol⁻¹) could accumulate in *Lib12.5*, whereas in *Lib250* only mutations exhibiting $\Delta\Delta G \leq 3.5$ kcal mol⁻¹ were tolerated (a higher $\Delta\Delta G$ value corresponds to more destabilizing mutations). However, by the tenth round, only weakly destabilizing mutations ($\Delta\Delta G \leq 3.5$ kcal mol⁻¹) accumulated in both populations (Fig. 3b).

The FoldX analysis therefore indicated the existence of a 'neutral' region in which changes in stability had only a mild effect on fitness, and, notably, that the width of this region changed with fitness level. As more mutations accumulated, this threshold was exhausted, and the tolerance to mutations under both regimes became essentially the same (Fig. 3b). This analysis is also in line with the finding that the percentage of tolerated mutations within *Lib12.5* was high at the beginning, and decreased during subsequent rounds (Table 1 and Supplementary Fig. 5). By the fifth round of *Lib12.5*, an average of 3.6 non-synonymous mutations per gene had accumulated, whereas only 1.9 mutations were added over the subsequent five rounds. In contrast, *Lib250* proceeded at a steady rate (2.2 non-synonymous

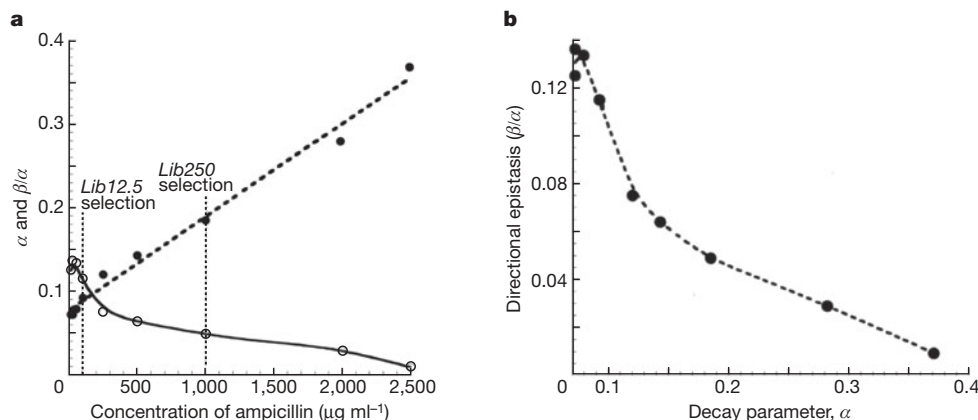


Figure 2 | The correlation between mutational robustness and negative epistasis. **a**, The decay parameter α (filled circles) and directional epistasis β/α (open circles) were extracted from the fitness decline fits of *Lib0* for each ampicillin concentration (Fig. 1, and Supplementary Fig. 4). These parameters are presented as a function of ampicillin concentration, corresponding to the fitness levels in our system. The vertical dotted lines

show the fitness levels for the purifying selections. **b**, The correlation between mutational robustness (higher α values indicate lower tolerance to mutations, and hence lower robustness) and epistasis. The same correlation was observed when an alternative measure of epistasis³ was applied (Supplementary Fig. 7).

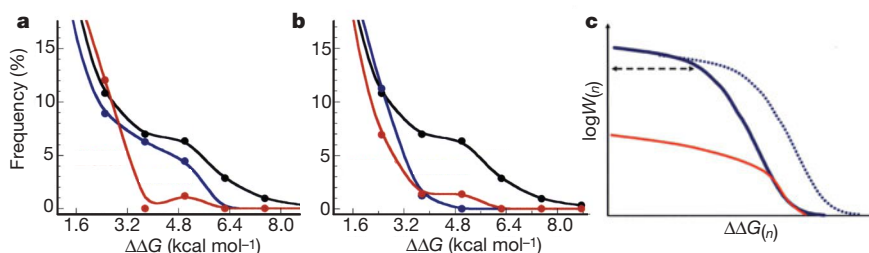


Figure 3 | The physico-chemical and fitness changes accompanying random drifts. **a, b,** The stability changes induced by 980 mutations identified in the three drifting populations were individually computed with FoldX. The calculated $\Delta\Delta G$ values were arranged in 1 kcal mol⁻¹ bins. Plotted are the frequencies of all destabilizing mutations found in the fifth round (**a**) and tenth round (**b**) for the unselected library (*Lib0*, black), and the two selected libraries (*Lib12.5* (blue) and *Lib250* (red)). **c,** Our model surmises that $\Delta\Delta G$ (physico-chemical changes relative to wild type, and in particular stability) declines linearly with the number of mutations

($\Delta\Delta G_{(n)} \propto n$). However, the decline in fitness $W_{(n)}$ that accompanies these $\Delta\Delta G$ changes is nonlinear and comprises both a threshold (dashed two-headed arrow) that buffers the deleterious effects of mutations, and a gradient in which fitness declines concomitantly with $\Delta\Delta G$ changes. At high fitness levels (solid blue line) the threshold is relatively narrow. At low fitness levels (red line) the threshold widens, giving rise to higher epistasis. A wider threshold is also predicted for protein variants carrying stabilizing mutations or global suppressors (dotted line; Supplementary Fig. 6).

mutations at the fifth round, and 4.6 at the tenth round). Indeed, the average numbers of mutations found in *Lib12.5* and *Lib250* by the tenth round of selection were nearly identical, despite a 20-fold difference in the stringency of selection.

It therefore seems that mutational robustness (or neutrality) should be described with two terms: 'threshold' and 'gradient' (Fig. 3c).

A 'threshold' induces a delay in fitness decline in the face of mutations. The threshold is observed not because the protein is unaffected by mutation but because the decline in its thermodynamic stability is buffered and has no immediate effect on fitness¹⁰. As mutations accumulate, however, the threshold is largely exhausted, and a gradient phase appears in which the fitness declines in parallel with $\Delta\Delta G$ changes. Because the robustness observed under lower fitness levels is largely due to a wider threshold, it is also inevitably epistatic (the fitness toll of later mutations is higher than that of the early ones). In fact, the manner in which α (robustness) and β (negative epistasis) correlate in our experiments (Fig. 2b) is in striking agreement with both theory¹¹ and simulations^{3–5}. The origins of our observed negative epistasis, and its correlation with robustness, therefore suggest that 'threshold robustness' is a general phenomenon that goes well beyond *TEM-1*.

Two conclusions can be derived from our results and model.

First, proteins may not be as robust as is generally assumed. So far, experiments have only tested the response to several random mutations and have reported an exponential decline in activity with no epistasis^{10,12,13}. The tolerance to mutations observed in these experiments is therefore likely to contain a major component of threshold robustness, especially as they were performed at low fitness^{10,14–16}. We, too, observed high levels of mutation tolerance under low fitness levels. Indeed, over the first five rounds of mutagenesis (70% of total mutations and 55% of non-synonymous mutations; Table 1), only 7% of mutations were unconditionally deleterious. These figures are

in agreement with the common view that the vast majority of protein mutations are tolerated^{10,14–17}. Yet once the stability threshold is exhausted, tolerance to mutations under the gradient regime is markedly lower and fits the view that most mutations affect fitness¹.

Second, our model provides general insights into the origins of robustness and epistasis, and why the two are interlinked^{3–5,11}. Threshold robustness relates to the margin of excess stability and function that could be compromised with little or no immediate effect on fitness, and is therefore inherently epistatic. Indeed, a quantitative correlation between the threshold (in $\Delta\Delta G$ terms) and directional epistasis (β/α) has been observed in our system and will be described in future work. Threshold robustness would be the type of robustness derived from redundancy at the genome level (for example duplicated genes)¹⁸, from stabilizing mutations¹⁹ or from global suppressors in single genes¹⁰ (Fig. 3c). It has been shown, for example, that *TEM-1* carrying a global suppressor mutation that increased stability by 2.7 kcal mol⁻¹ shows higher robustness relative to the wild type¹⁰ ($\alpha = 0.042$ versus 0.152 for wild type, by equation (1)). However, when we applied our model (equation (2)), we also recorded a much higher epistasis ($\beta/\alpha = 0.598$ versus 0.144; Supplementary Fig. 6).

Thus, theory and simulations^{3–5} have predicted a tight correlation between robustness and epistasis. Our work provides an experimental verification of this correlation and proposes a mechanism that accounts for it. Our model implies that any biological system that exhibits threshold robustness, or redundancy robustness¹⁸, is inevitably epistatic. In such systems, mechanisms that purge potentially deleterious mutations, such as recombination (through sexual reproduction and other mechanisms^{20,21}) are of crucial importance, as they help to maintain this threshold. In this way, recombination, threshold robustness and negative epistasis may be interlinked—each being an inevitable by-product of the other⁴.

Table 1 | Sequence data of the *TEM-1* libraries

Library	Round of mutagenesis and selection	No. of variants (base pairs) sequenced	Total no. of mutations detected	Average no. of mutations per gene	Average no. of non-synonymous mutations per gene	Percentage of tolerated mutations	
						Total	Non-synonymous
<i>Lib0</i>	1	10 (8,610)	16	1.6 ± 1.4	1.11 ± 1.0		
	2	21 (18,081)	81	3.9 ± 1.6	2.67 ± 1.1		
	3	5 (4,305)	25	5.0 ± 0.7	3.46 ± 0.5		
	5	17 (14,637)	161	9.5 ± 3.7	6.55 ± 2.6		
	10	14 (12,054)	290	20.7 ± 2.8	14.3 ± 1.9		
<i>Lib12.5</i>	2	16 (13,776)	52	3.3 ± 1.6	1.6 ± 1.2	84%	61%
	5	39 (33,579)	255	6.6 ± 3.0	3.6 ± 1.7	70%	55%
	10	17 (14,637)	171	10.7 ± 3.5	5.5 ± 2.2	52%	39%
<i>Lib250</i>	2	14 (12,054)	24	1.7 ± 1.2	1.3 ± 0.9	44%	48%
	5	39 (33,579)	193	5.0 ± 2.2	2.2 ± 1.5	52%	33%
	10	18 (15,498)	170	9.4 ± 3.0	4.6 ± 2.2	46%	32%

Where errors are shown, results are means ± s.d.

METHODS

Detailed methods are provided in Supplementary Information.

Library construction and selection. The *TEM-1* β -lactamase open reading frame was recloned (*NcoI/NotI*) under the control of its endogenous promoter into a modified pUC19 plasmid containing a chloramphenicol resistance gene. Random mutagenesis was performed by 'wobble'-base polymerase chain reaction with *TEM-1* plasmids as template. The protocol was optimized to achieve an average of two mutations per gene per generation, and a characteristic pattern of nucleotide exchange (Supplementary Table 2). Plasmid libraries were obtained by electroporation into *E. coli* and more than 10^6 individual transformants were grown in Luria–Bertani medium in the presence of chloramphenicol. DNA extracted from these cultures was retransformed into XL-1 Blue *E. coli* cells and plated on agar plates supplemented with ampicillin (0, 12.5 or 250 $\mu\text{g ml}^{-1}$). Plasmid DNA from surviving colonies was extracted and served as a template for the subsequent round of mutagenesis.

Fitness measurements. Several hundred colonies from each library were randomly picked and grown in 96-well plates in the presence of chloramphenicol. The 96-well plates were replica-plated onto agar plates with the entire range of ampicillin concentrations (0–2,500 $\mu\text{g ml}^{-1}$). Fitness values (*W*) were determined from the fraction of clones that survived each ampicillin concentration (two to five independent repetitions of this protocol were performed and standard deviations are presented). The reproducibility of the mutagenesis protocol was verified by an independent repetition of one round of mutagenesis, indicating a variability within the deviation of the fitness measurements (Supplementary Table 1).

Computation of stability effects. FoldX is a structure-based algorithm that predicts the $\Delta\Delta G$ values of mutations with considerable fidelity ($R = 0.83$ for a data set of 1,030 mutations in 27 proteins)⁶. We applied the published procedure, including an adjustment of the calculated values²². A correlation between the deleterious effect of mutations and their predicted $\Delta\Delta G$ values was also obtained (Supplementary Table 3).

Received 1 September; accepted 27 October 2006.

Published online 19 November 2006.

1. Pal, C., Papp, B. & Lercher, M. J. An integrated view of protein evolution. *Nature Rev. Genet.* **7**, 337–348 (2006).
2. DePristo, M. A., Weinreich, D. M. & Hartl, D. L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nature Rev. Genet.* **6**, 678–687 (2005).
3. Wilke, C. O. & Adami, C. Interaction between directional epistasis and average mutational effects. *Proc. R. Soc. Lond. B* **268**, 1469–1474 (2001).
4. Azevedo, R. B., Lohaus, R., Srinivasan, S., Dang, K. K. & Burch, C. L. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature* **440**, 87–90 (2006).
5. Elena, S. F., Carrasco, P., Daros, J. A. & Sanjuan, R. Mechanisms of genetic robustness in RNA viruses. *EMBO Rep.* **7**, 168–173 (2006).
6. Guerois, R., Nielsen, J. E. & Serrano, L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* **320**, 369–387 (2002).

7. Wolf, J. B., Brodie, E. D. & Wade, M. J. *Epistasis and the evolutionary process* (Oxford Univ. Press, Oxford, 2000).
8. Horovitz, A. Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Fold. Des.* **1**, R121–R126 (1996).
9. Charlesworth, B. Mutation–selection balance and the evolutionary advantage of sex and recombination. *Genet. Res.* **55**, 199–221 (1990).
10. Bloom, J. D. et al. Thermodynamic prediction of protein neutrality. *Proc. Natl Acad. Sci. USA* **102**, 606–611 (2005).
11. Wagner, G. P., Laubichler, M. D. & Bagheri-Chaichian, H. Genetic measurement of theory of epistatic effects. *Genetica* **102–103**, 569–580 (1998).
12. Shafikhani, S., Siegel, R. A., Ferrari, E. & Schellenberger, V. Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. *Biotechniques* **23**, 304–310 (1997).
13. Daugherty, P. S., Chen, G., Iverson, B. L. & Georgiou, G. Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. *Proc. Natl Acad. Sci. USA* **97**, 2029–2034 (2000).
14. Guo, H. H., Choe, J. & Loeb, L. A. Protein tolerance to random amino acid change. *Proc. Natl Acad. Sci. USA* **101**, 9205–9210 (2004).
15. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. & Sauer, R. T. Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* **247**, 1306–1310 (1990).
16. Sauer, U. H., San, D. P. & Matthews, B. W. Tolerance of T4 lysozyme to proline substitutions within the long interdomain α -helix illustrates the adaptability of proteins to potentially destabilizing lesions. *J. Biol. Chem.* **267**, 2393–2399 (1992).
17. Huang, W., Petrosino, J., Hirsch, M., Shenkin, P. S. & Palzkill, T. Amino acid sequence determinants of β -lactamase structure and activity. *J. Mol. Biol.* **258**, 688–703 (1996).
18. Wagner, A. Distributed robustness versus redundancy as causes of mutational robustness. *BioEssays* **27**, 176–188 (2005).
19. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. A. Protein stability promotes evolvability. *Proc. Natl Acad. Sci. USA* **103**, 5869–5874 (2006).
20. Kondrashov, A. S. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440 (1988).
21. Lynch, M., Burger, R., Butcher, D. & Gabriel, W. The mutational meltdown in asexual populations. *J. Hered.* **84**, 339–344 (1993).
22. Kiel, C. & Serrano, L. The ubiquitin domain superfold: structure-based sequence alignments and characterization of binding epitopes. *J. Mol. Biol.* **355**, 821–844 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F. Kondrashov for insights and inspiration, L. Serrano, U. Alon and G. Schreiber for contributions, and our Department, Institute and the Feinberg Graduate School, for uncompromising support.

Author Contributions S.B. designed, performed and analysed the experiments. M.S. and R.B. provided technical assistance. N.T. performed the FoldX computations and helped devise the model. D.S.T. designed and supervised the experiments, and devised the model. S.B. and D.S.T. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.S.T. (tawfik@weizmann.ac.il).

Characterization of a carbohydrate transporter from symbiotic glomeromycotan fungi

Arthur Schüßler^{1†}, Holger Martin¹, David Cohen¹, Michael Fitz² & Daniel Wipf²

The symbiotic relationships between mycorrhizal fungi and plants have an enormous impact on terrestrial ecosystems¹. Most common are the arbuscular mycorrhizas, formed by fungi belonging to the phylum Glomeromycota². Arbuscular mycorrhizal fungi facilitate the uptake of soil nutrients by plants³ and in exchange obtain carbohydrates, thus representing a large sink⁴ for atmospheric plant-fixed CO₂. However, how carbohydrates are transported through the symbiotic interface is still unknown. Here we report the characterization of the first known glomeromycotan monosaccharide transporter, GpMST1, by exploiting the unique symbiosis of a glomeromycotan fungus (*Geosiphon pyriformis*) with cyanobacteria⁵. The *GpMST1* gene has a very low GC content and contains six introns with unusual boundaries. GpMST1 possesses twelve predicted transmembrane domains and functions as a proton co-transporter with highest affinity for glucose, then mannose, galactose and fructose. It belongs to an as yet uncharacterized phylogenetic monosaccharide transporter clade. This initial characterization of a new transporter family involved in fungal symbiosis will lead to a better understanding of carbon flows in terrestrial environments.

Members of the Glomeromycota are multikaryotic, asexual, obligate symbionts, and as such render certain studies difficult or even impossible. We use the unique symbiosis of a glomeromycotan fungus (*G. pyriformis*) with a cyanobacterium (*Nostoc punctiforme*) to get a handle on the role of fungal transcripts in carbohydrate transport⁴. This symbiosis is highly interesting because (1) it is the only known fungal endosymbiosis with cyanobacteria and (2) the symbiotic stage (that is, the 'bladder') is comparable to the within-root situation of the arbuscular mycorrhiza^{5,6} (Fig. 1). The bonus for gene expression studies is that, contrary to 'real' arbuscular mycorrhizas, fungal messenger RNA (mRNA) can be isolated specifically from the symbiotic stage.

To characterize the first glomeromycotan sugar transporter we functionally complemented the hexose-uptake deficient yeast strain⁷ EBY.VW4000 with a complementary DNA (cDNA) expression library established from symbiotic *G. pyriformis* bladders. Screenings resulted in yeast clones containing plasmids with a cDNA insert of 2,043 bp (Fig. 2a). The cDNA encodes an ORF of 540 amino acids, which contains the PROSITE sugar transport proteins signature 2, PS00217 (<http://www.expasy.org/prosite/PS00217>; consensus pattern [LIVMF]xG[LIVMFA]{V}xG[KP]-x(7)-[LIFY]-x(2)-[EQ]-x(6)-[RK]), and the sugar transporter PFAM motif, PF00083 (<http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00083>). The gene was shown (see below) to code for a monosaccharide transporter (MST) and named *GpMST1*. It codes for a 12 transmembrane domain (TMD) topology⁸ transporter of the major facilitator superfamily with intracellular termini, a large extracellular loop (between TMD1 and TMD2) that is shared by most hexose transporters⁹, and a relatively long¹⁰ central loop (Fig. 2b).

Phylogenetic analyses were performed to discover to which of the known MST clades *GpMST1* belongs (Supplementary Figs 1, 2). Results of BLASTP searches comprised only 10–14 fungal sequences within the 200 best hits, with highest identities/similarities of 40/60%, respectively, to animal and fungal sequences. Phylogenetic analyses of 1,620 transporters revealed that *GpMST1* does not cluster within any of the previously defined MST clades¹⁰, but within a distant fungal clade. The clade consists of (mostly from recent genome projects) sequences from 16 ascomycete fungi (*Saccharomyces cerevisiae* Q6B2Y5; YB91_YEAST; VPS73_YEAST; *Kluyveromyces lactis* Q8J289; *Candida glabrata* Q6FWZ2; *Candida albicans* Q59UD9 and Q5ALJ1; *Debaryomyces hansenii* Q6BY36 and Q6BJ92; *Yarrowia lipolytica* Q6CDU0; *Ashbya gossypii* Q754V5; *Aspergillus fumigatus* Q4WYR5; *Aspergillus oryzae* Q2UC56; *Gibberella zeae* Q4IE71; *Blumeria graminis* Q8NK49; and *Neurospora crassa* Q870X7) and 3 basidiomycete fungi (*Cryptococcus neoformans* Q55TK8 and Q55VM4; and *Ustilago maydis* Q4PE10). *GpMST1* is presumably the first characterized transporter of this new and interesting clade of fungal MSTs. However, because the phylogenetic distance between dikaryotic fungi (Ascomycota and Basidiomycota) and Glomeromycota is very large (putative split >700 Myr ago) the clade is not strongly supported and more glomeromycotan sequences will be needed before drawing any final conclusions.

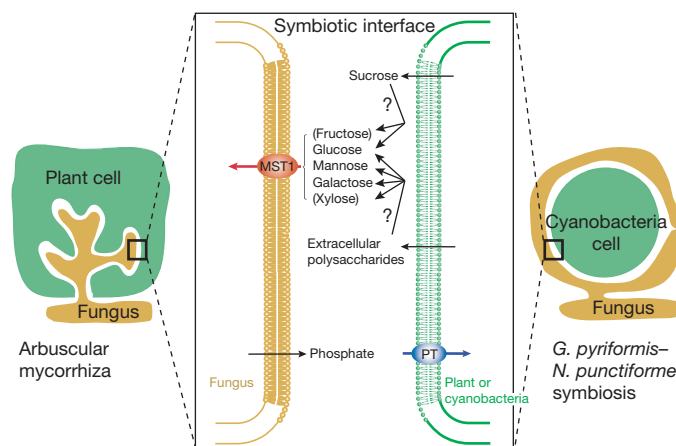


Figure 1 | Comparison of bidirectional phosphate and carbon exchanges. Despite inverted relative dimensions of macro- and micro-symbiont the interface and nutrient exchange in the *G. pyriformis* symbiosis correspond to that in the arbuscular mycorrhiza. Several arbuscular mycorrhiza-specific phosphate transporters (PT) are known from plants. The hypothetical role of *GpMST1*, and its orthologues, in the sugar uptake through the symbiotic membrane of glomeromycotan fungi is indicated together with the substrates of *GpMST1* (fructose and putatively xylose are transported weakly).

¹Darmstadt University of Technology, Institute of Botany, Schnittspahnstrasse 10, 64287 Darmstadt, Germany. ²University of Bonn, IZMB, NWG Transport in Mycorrhiza, Kirschallee 1, 53115 Bonn, Germany. [†]Present address: Bereich Genetik, Department Biologie I, Maria-Ward-Straße 1a, 80638 München, Germany.

The *GpMST1* gene is 2,678 bp long and contains 6 introns (Fig. 2a) and 3 functional polyadenylation signals (cDNA with poly(A) tails beginning after all three signals were sequenced). A typical feature of arbuscular mycorrhizal fungi (AMF) genomes seems to be an exceptionally low GC content. For example, *Glomus intraradices* has an average of 36.3% GC among its codons¹¹, and the same holds true for *GpMST1* (codons 37.6%; genomic DNA 31.4%; introns 22%). The intron–exon boundaries of the six introns (Supplementary Fig. 3) are significantly influenced by the low GC content. Half of them contain an A as the first exon-nucleotide at the 3' splicing site, and, unusually, two have the pyrimidine T. At the third position upstream of the 3' splicing site >90% of human and yeast¹² introns contain a pyrimidine (T or C), but two-thirds of the *GpMST1* introns show a purine (A) base. Such unusual features may be expected for other glomeromycotan genes, too. To reveal whether *GpMST1* is expressed in a photosynthesis-dependent manner, its expression after 47 or 70 h of darkness, or after 18, 47 or 70 h of light was studied by semi-quantitative RT–PCR (PCR after reverse transcription): the gene seems to be constitutively expressed over this time (data not shown).

To investigate the substrate specificity, the *GpMST1* expressing yeast mutant EBY.VW4000 was plated on sugars and sugar derivatives. The clones could not grow on D-sucrose, D-lactose, D-cellobiose, D-trehalose, D-arabinose, L-arabinose, D-ribose, D-xylose, myo-inositol, D-sorbitol, D-glucosamine N-acetyl-glucosamin, L-glucose, L-rhamnose and L-fucose. Growth was reconstituted on D-fructose (very weak), D-galactose (moderate), D-mannose (high) and D-glucose (highest). To investigate whether growth rates were due to the affinity of the transporter or to altered yeast metabolism, competition experiments were performed (Fig. 3a). Xylose was also tested because it is a main constituent of plant cell walls and some fungal MSTs transport xylose at low rates^{13,14}; indeed, this was also indicated for *GpMST1*. Uptake of ¹⁴C-glucose (at pH 6.5) through *GpMST1* was very sensitive to protonophores and H⁺-ATPase inhibitors (Fig. 3b), indicating secondary active H⁺ co-transport. The glucose uptake was pH-dependent, with a somewhat surprising optimum of pH 7 (Fig. 3c). We used pH 6.5 for all experiments to ensure working with a $\Delta pH > 0.5$, and because most yeast sugar

transport studies were performed at this pH. Michaelis–Menten kinetics and Lineweaver–Burk transformation indicate a $V_{\max}$ of ~0.18 nmol glucose per 10⁶ cells per min and a K_M of ~1.2 mM for glucose uptake (Fig. 3d). *GpMST1* therefore represents the first known glomeromycotan sugar transporter, transporting monosaccharides with rates in the order glucose>mannose>galactose>fructose.

Many symbiotic *Nostoc* strains seem to be capable of heterotrophic growth on sucrose, glucose or fructose, such as the strain PCC 73102—which forms symbioses with *Anthoceros* and *Gunnera*¹⁵—and also the symbiont of *G. pyriformis*¹⁶. However, contrary to the plant symbioses formed with *Anthoceros* and *Gunnera*, carbohydrate transport in the *G. pyriformis* symbiosis is in the opposite direction, from photoautotrophic symbiont to fungus, as in the arbuscular mycorrhiza.

There are some indications that *GpMST1* transports hexoses across the symbiotic interface membrane. On the one hand, when expressed in yeast, it is clearly active in the plasma membrane; on the other hand, for *G. pyriformis* incubated in ¹⁴C-glucose no uptake from the outside could be shown (M. Kluge, personal communication), although metabolites originating from cyanobacterial ¹⁴CO₂ fixation could be easily detected¹⁷. The *G. pyriformis* membrane at the symbiotic interface is derived from the plasma membrane (Fig. 1), and a chitin-containing 'cell wall' layer, which is similar to the arbuscule cell wall, is synthesized within the bladders¹⁸. Our interpretation is that *GpMST1* probably represents the type of MST that is responsible for the transport of plant carbohydrates to glomeromycotan fungi through the symbiotic membrane, and that *G. pyriformis*, as well as 'typical' AMF^{4,19}, are not able to take up glucose through the non-symbiotic plasma membranes.

Carbon flow is a key process in the arbuscular mycorrhiza, and sucrose is usually interpreted to be the source of hexoses taken up by AMF. The *N. punctiforme* symbiont of *G. pyriformis* synthesizes sucrose¹⁵ for osmoregulation, which therefore could also be a C source for the fungus. However, another aspect might be interesting: besides glucose, *GpMST1* efficiently transports main 'cell wall hexoses'—mannose and galactose. This unusual substrate specificity led us to

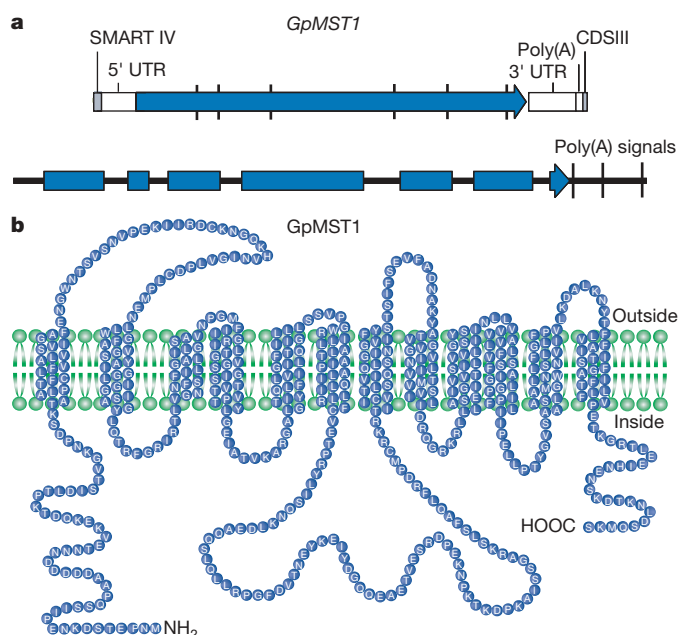


Figure 2 | The *G. pyriformis* MST1. **a**, The *GpMST1* gene contains six introns. The splicing sites as well as the UTRs are indicated; the SMART and CDS oligonucleotides at the cDNA termini (introduced by the cDNA library construction system used) carry the *Sfi* I restriction sites used for cloning; three poly(A)-signals were shown to be functional. **b**, The amino acid sequence and putative topology of *GpMST1* are shown.

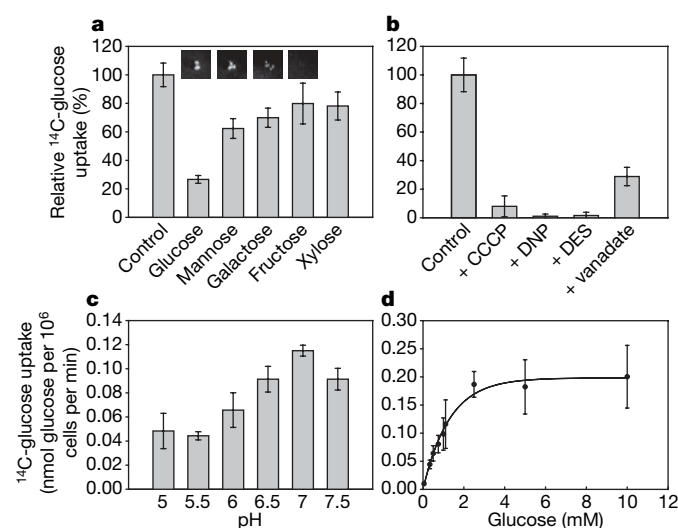


Figure 3 | ¹⁴C-glucose uptake in *GpMST1*-expressing yeast strain EBY.VW4000. **a**, Substrate competition and colony growth (inserts) on the respective sugars; *GpMST1* reconstitutes growth on glucose, mannose, galactose, and (very weakly) fructose; xylose is indicated to be transported but cannot be metabolized. **b**, Incubation in the presence of protonophores (CCCP, DNP) and plasma membrane H⁺-ATPase inhibitors (DES, vanadate) for 5 min strongly inhibits glucose uptake. **c**, The pH optimum for uptake is about pH 7, in the yeast system. **d**, Michaelis–Menten kinetics of glucose uptake rates (pH 6.5) indicate a K_M of ~1.2 mM. Error bars represent s.d.; $n = 3$.

the speculation that at least a portion of the monosaccharides could be liberated from plant extracellular polysaccharides and taken up as a C source by AMF, as indicated in Fig. 1. In the *G. pyriformis* symbiosis the extracellular polysaccharides synthesized by *N. punctiforme* seem to be degraded. In arbuscular mycorrhizas, the plant 'cell wall' in the interface space is also very thin and was shown to be amorphous and to contain mainly glucose, mannose, galactose, xylose, and arabinose²⁰. Recently, high activities of xyloglucanases and endocellulases were reported in arbuscular mycorrhizal roots^{21,22}, and it was shown that the growth of AMF is significantly enhanced by certain sugars that are not derived from sucrose cleavage, for example, raffinose and other trisaccharides²³. Might extracellular polysaccharides be an alternative energy source for AMF? Could the plant vesicle transport towards the fast-growing perisymbiotic membrane be involved in C-supply and regulation of root colonisation?

There are many open questions, but the glomeromycotan MST characterized here should boost this field by enabling the isolation and characterization of orthologues, which will allow expression profiling in arbuscular mycorrhizas. This may lead, in the near future, to a much better understanding of arbuscular mycorrhizas and the related global carbon fluxes.

METHODS

Organisms. The *G. pyriformis*–*N. punctiforme* symbiosis was grown in our laboratory as described⁵. *S. cerevisiae* strains used were the *hxt*-null mutant EBY.VW4000 and its parental strain CEN.PK2-1C (ref. 7).

Messenger RNA isolation, reverse transcription and cDNA amplification. Five-hundred *G. pyriformis* bladders were frozen in liquid N₂ within a 2 ml vial containing a tungsten carbide bead. Deep-frozen samples were ground in a bead mill (MM2000, Retsch) and ~80 ng of poly(A)-mRNA was isolated with magnetic oligo(dT)-beads (Dynabeads mRNA DIRECT kit, Dynal Biotech). Reverse transcription was performed and double-stranded cDNA amplified with the Creator SMART ds-cDNA synthesis kit (Clontech, BD Biosciences).

Construction of the cDNA library. Amplified cDNA was size-fractionated (>1.1 kb) and concentrated. Five-hundred ng was digested with *Sfi* I and ligated into *pDR196* vector²⁴, which was modified to contain asymmetric *Sfi* I sites. *pDR196sfi*-cDNA plasmids were then agarose-gel-extracted and transfected into *Escherichia coli* XL 10 Gold (Stratagene). Clones (6×10^5) were harvested; ~95% were recombinant and average insert size was ~1.1 kb.

Isolation of *GpMST1* by functional complementation. Clones (10^5) were harvested from the primary library. After transformation²⁵ of EBY.VW4000, selection was performed on synthetic dextrose (SD) agar (without uracil) containing 2% maltose instead of glucose. Colonies were harvested and aliquots plated on SD agar (without uracil) containing 22 mM glucose as the C source. *pHM13-C6* (representative of several clones) was isolated, transfected into *E. coli*, re-isolated, and re-transfected into EBY.VW4000 for further characterization of three sub-clones.

Genomic DNA sequencing and amplification. Genomic DNA (1 ng) from *G. pyriformis* was amplified with the Phi29-DNA- polymerase-based GenomiPhi DNA amplification kit (GE Healthcare), resulting in >5 µg of amplified gDNA. Primers designed for the termini of cDNA sequences were used to amplify the *GpMST1* gene with the Phusion proof-reading polymerase (Finnzymes). The full-length fragment (*pDC13-C6*) was cloned in *pCR4Blunt-TOPO* (Invitrogen) and sequenced, as well as a further eight overlapping fragments from three independent gDNA extractions.

Semi-quantitative RT-PCR. *G. pyriformis* bladders were stochastically distributed (5 Petri dishes, 50 bladders each) and subjected to 18, 47 or 70 h of light, or 47 or 70 h darkness. Bladders were collected in RNase-free vials, immediately frozen in liquid N₂, and mRNA was isolated using Dynabeads, but without heat elution. The beads-mRNA complex was resuspended in reverse transcription mixture (Sensiscript RT kit, QIAGEN, Hilden). Reverse transcription was performed using the oligo(dT)-beads as primer, resulting in covalently attached cDNA. A 720 bp *GpMST1* and a 520 bp *EF1α* cDNA fragment (specificity checked by sequencing) were sequentially amplified by solid-phase PCR with Phusion proof-reading polymerase. Samples were taken after 15, 19, 23, 27 and 31 cycles and analysed by agarose gel electrophoresis.

¹⁴C-glucose uptake studies. Yeast cells were grown to A_{600 nm} 0.5, harvested, washed in water twice and resuspended in buffer A (0.6 M sorbitol, 50 mM potassium phosphate) to A_{600 nm} 5.0. Before measurements, cells were incubated in 8 mM maltose for 5 min. The reaction was started by adding 100 µl of cell suspension to 100 µl of buffer A containing at least 7.4 kBq of ¹⁴C-glucose (Amersham Biosciences) and unlabelled glucose at the concentrations used in

the experiments. After 30, 60, 120 and 240 s, 50 µl aliquots were transferred to 4 ml of ice-cold buffer A, filtered on glass fibre filters, and washed twice with buffer before liquid scintillation spectrometry. Competition for glucose uptake was performed by adding a fivefold molar excess (5.5 mM) of the respective competitors to 1.1 mM glucose. pH dependence was analysed in 100 mM potassium phosphate buffer adjusted to different pH values. The influence of the electrochemical gradient was analysed by incubation (5 min) in 1.1 mM glucose (control), or in glucose and 2,4-dinitrophenol (DNP), diethylstilbestrol (DES), carbonyl cyanide *m*-chlorophenylhydrazone (CCCP) or vanadate (100 µM each). All measurements were repeated independently in at least three experiments.

Phylogenetic analysis. The *GpMST1* amino acid sequence was first aligned with the 200 best BLASTP hits. From the alignment, 251 sites could be used. Identical sequences were removed and 170 sequences analysed with a quartet puzzling maximum likelihood method (TREE-PUZZLE 5.2). Frequencies of amino acids were estimated from the data set, rate heterogeneity taken into account (invariable and gamma-distributed rates), and WAG, JTT and VT models used. Further analyses used the transporter alignment (2,174 sequences) from <http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00083>. After removing short sequences the data set was composed of 1,620 sequences (375 sites) and the fungal data set of 742 sequences (380 sites). The 1,620 sequences were used to construct distance and parsimony trees with PROTDIST and PROTPARS (PHYLIP 3.6; see Supplementary Figs 1 and 2).

Received 9 July; accepted 20 October 2006.

- Smith, S. E. & Read, D. J. *Mycorrhizal Symbiosis* (Academic Press, London, 1997).
- Schüßler, A., Schwarzott, D. & Walker, C. A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycol. Res.* **105**, 1413–1421 (2001).
- Marx, J. The roots of plant-microbe collaborations. *Science* **304**, 234–236 (2004).
- Bago, B. et al. Carbon export from arbuscular mycorrhizal roots involves the translocation of carbohydrate as well as lipid. *Plant Physiol.* **131**, 1496–1507 (2003).
- Schüßler, A. & Wolf, E. *Geosiphon pyriformis*—a glomeromycotan soil fungus forming endosymbiosis with cyanobacteria. In *In vitro culture of mycorrhizas* (Soil Biology, Volume 4) (eds Declerck, S., Strullu, D.-G., Fortin, J. A.) 271–289 (Springer, Berlin Heidelberg, 2005).
- Schüßler, A. Molecular phylogeny, taxonomy, and evolution of *Geosiphon pyriformis* and arbuscular mycorrhizal fungi. *Plant Soil* **244**, 75–83 (2002).
- Wieczorke, R. et al. Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. *FEBS Lett.* **464**, 123–128 (1999).
- Sonnhammer, E. L. L., von Heijne, G. & Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. In *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology (ISMB98)* (ed. Glasgow, J., et al.) 175–182 (AAAI Press, Menlo Park, 1998).
- Gould, G. W. & Seater, M. J. Introduction to the facilitative glucose transporter family. In *Facilitative glucose transporters (Molecular Biology Intelligence Unit)* (ed. Gould, G. W.) 1–38 (Springer, Berlin Heidelberg, 1997).
- Lalonde, S., Wipf, D. & Frommer, W. B. Transport mechanisms for organic forms of carbon and nitrogen between source and sink. *Annu. Rev. Plant Biol.* **55**, 341–372 (2004).
- Sawaki, H. & Saito, M. Expressed genes in the extraradical hyphae of an arbuscular mycorrhizal fungus, *Glomus intraradices*, in the symbiotic phase. *FEMS Microbiol. Lett.* **195**, 109–113 (2001).
- Spingola, M., Grate, L., Haussler, D. & Ares, M. Jr. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* **5**, 221–234 (1999).
- Hamacher, T., Becker, J., Gardonyi, M., Hahn-Hägerdal, B. & Boles, E. Characterization of the xylose-transporting properties of yeast hexose transporters and their influence on xylose utilization. *Microbiology* **148**, 2783–2788 (2002).
- van Kuyk, P. A. et al. *Aspergillus niger mstA* encodes a high-affinity sugar/H⁺ symporter which is regulated in response to extracellular pH. *Biochem. J.* **379**, 375–383 (2004).
- Meeks, J. C. et al. An overview of the genome of *Nostoc punctiforme*, a multicellular, symbiotic cyanobacterium. *Photosynth. Res.* **70**, 85–106 (2001).
- Knoblauch, M. Vergleichende physiologische Charakterisierung von *Nostoc punctiforme* und *Nostoc muscorum* im Hinblick auf die Symbiose von *Nostoc* in *Geosiphon pyriforme*. Master's Thesis, 1–70, Univ. of Frankfurt (1995).
- Kluge, M., Mollenhauer, D. & Mollenhauer, R. Photosynthetic carbon assimilation in *Geosiphon pyriforme* (Kützinger) F.v. Wettstein, an endosymbiotic association of fungus and cyanobacterium. *Planta* **185**, 311–315 (1991).
- Schüßler, A., Bonfante, P., Schnepf, E., Mollenhauer, D. & Kluge, M. Characterization of the *Geosiphon pyriforme* symbiosome by affinity techniques: confocal laser scanning microscopy (CLSM) and electron microscopy. *Protoplasma* **190**, 53–67 (1996).
- Solaiman, Z. & Saito, M. Use of sugars by intraradical hyphae of arbuscular mycorrhizal fungi revealed by radiorespirometry. *New Phytol.* **136**, 533–538 (1997).

20. Bonfante, P. At the interface between mycorrhizal fungi and plants: the organization of cell wall, plasma membrane and cytoskeleton. In *The Mycota IX—Fungal Associations* (ed. Hock, B.) 45–61 (Springer, Berlin Heidelberg New York, 2001).
21. García, M., Sampedro, I., Ocampo, J. A. & García-Romera, I. Xyloglucanases in the interaction between *Sinorhizobium meliloti*, *Rhizobium leguminosarum* bv. *viciae* and *Glomus mosseae*. In *Abstracts of the 5th International Conference on Mycorrhizas, 23–27 July 2006—Granada, Spain* (eds Barea, J. M., Azcón, C., Gutiérrez, F., González, F., Molina, A. J.) 200 (Gráficas Zaidín, Granada, 2006).
22. Morales Vela, G., Molinero-Rosales, N., Algaba, N., Ocampo Bote, J. A. & García Garrido, J. M. Endocellulase activity in pea symbiotic mutants colonized by arbuscular mycorrhizal fungi. In *Abstracts of the 5th International Conference on Mycorrhizas, 23–27 July 2006—Granada, Spain* (eds Barea, J. M., Azcón, C., Gutiérrez, F., González, F., Molina, A. J.) 40 (Gráficas Zaidín, Granada, 2006).
23. Hildebrandt, U., Ouziad, F., Marnier, F.-J. & Bothe, H. The bacterium *Paenibacillus validus* stimulates growth of the arbuscular mycorrhizal fungus *Glomus intraradices* up to the formation of fertile spores. *FEMS Microbiol. Lett.* **254**, 258–267 (2006).
24. Wipf, D. *et al.* An expression cDNA library for suppression cloning in yeast mutants, complementation of a yeast *his4* mutant and EST analysis from the symbiotic basidiomycete *Hebeloma cylindrosporum*. *Genome* **46**, 177–181 (2003).
25. Gietz, D., St Jean, A., Woods, R. A. & Schiestl, R. H. Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res.* **20**, 1425 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Boles (University of Frankfurt) for providing EBY.VW4000 and CEN.PK2-1C yeast strains, D. Rentsch (University of Bern) for providing *pDR196*, and J. Gossmann (University of Frankfurt) for proofreading. This work was supported by grants from the Deutsche Forschungsgemeinschaft to D.C., H.M. and A.S.

Author Contributions A.S., H.M. and D.C. isolated the full-length *GpMST1* cDNA clones and performed semiquantitative RT–PCR. A.S. and D.C. retransformed EBY.VW4000, and performed screenings and growth tests. H.M. modified the *pDR196* vector, constructed the cDNA libraries, and performed initial yeast transformations and screenings. M.F. performed the uptake assays. D.C. helped with uptake assays and performed gDNA amplification, PCR and clonings. A.S. made the phylogenetic analyses; A.S. and D.W. are responsible for the experimental design, hypotheses, interpreting the results and writing the manuscript.

Author Information Sequences are deposited at EMBL Data Bank with the accession numbers AM231332 (*GpMST1* cDNA clone *pHM13-C6.1*) and AM231333 (*GpMST1* gDNA clone *pDC-C6*). Plasmids and clones are available on request. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.S. (arthur.schuessler@lrz.uni-muenchen.de).

The circumsporozoite protein is an immunodominant protective antigen in irradiated sporozoites

Kota Arun Kumar^{1*}, Gen-ichiro Sano^{2*}, Silvia Boscardin³, Ruth S. Nussenzweig⁴, Michel C. Nussenzweig³, Fidel Zavala⁵ & Victor Nussenzweig¹

Malaria infection starts when mosquitoes inject sporozoites into the skin. The parasites enter the blood stream and make their way to the liver where they develop into the exo-erythrocytic forms (EEFs). Immunization with irradiated sporozoites (IrSp) leads to robust protection against malaria infection in rodents¹, monkeys² and humans³ by eliciting antibodies to circumsporozoite protein (CS) that inhibit sporozoite infectivity, and T cells that destroy the EEFs⁴. To study the role of non-CS antigens in protection, we produced CS transgenic mice that were tolerant to CS T-cell epitopes. Here we show that in the absence of T-cell-dependent immune responses to CS, protection induced by immunization with two doses of IrSp was greatly reduced. Thus, although hundreds of other *Plasmodium* genes are expressed in sporozoites⁵ and EEFs⁶, CS is a dominant protective antigen. Nevertheless, sterile immunity could be obtained by immunization of CS transgenics with three doses of IrSp.

As shown in the late 1960s, immunization of mice with γ -irradiated rodent malaria sporozoites elicited complete protection against sporozoite challenge¹. These findings were extended to humans who were vaccinated through the bites of several hundred irradiated, *Plasmodium falciparum*-infected mosquitoes³. The human trials showed that protection was species- but not isolate-specific⁷, paving the way for the development of pre-erythrocytic human malaria vaccines. Sporozoites and EEFs contain many potential antigens, and it has been argued that IrSp immunization is successful because it elicits immune responses that are broad-based against multiple antigenic targets⁸. If correct, this implies that pre-erythrocytic vaccines will be effective only if they contain attenuated sporozoites, or large numbers of antigens⁹. An alternative hypothesis is that a few immunodominant antigens mediate the protection generated by IrSp.

To evaluate the contribution of CS to protection by immunization with *Plasmodium yoelii* IrSp, we produced transgenic mice expressing CS under the control of the *Gt(ROSA)26Sor* (ROSA26) promoter (Fig. 1a). The transgenic mice were identified by PCR screening of genomic DNA (Fig. 1b). As shown in Fig. 1c, CS was detected in the blood of the transgenics, and CS messenger RNA was expressed in several organs (Fig. 1d). Circulating CS inhibited the infectivity of live sporozoites injected either intravenously or delivered by mosquito bite. The inhibition was inversely related to the dose of injected parasites (Fig. 2a–d), and was almost complete when sporozoites were injected by mosquitoes (Fig. 2e). These findings indicate that the circulating CS protein inhibits infection by competing for the ligand on the hepatocyte surface, consistent with several earlier *in vitro* observations^{10–12}.

To determine whether CS-transgene expression produces T-cell tolerance, transgenic BALB/c mice were primed with a recombinant yellow fever 17D vaccine expressing the H-2K^d restricted cytotoxic T lymphocyte (CTL) epitope of the *P. yoelii* CS, and boosted with a recombinant modified vaccinia Ankara virus expressing the entire coding sequence of CS¹³. We found that control littermates and conventional BALB/c mice developed hundreds of interferon- γ (IFN- γ) secreting T cells in response to immunization, whereas the CS transgenics did not (Fig. 3).

To examine the humoral response, CS transgenics, control littermates and BALB/c mice were immunized and boosted with 2×10^4 *P. yoelii* IrSp. We found specific antibodies against the major immunodominant repeats of CS in all mice. However, in the CS transgenics,

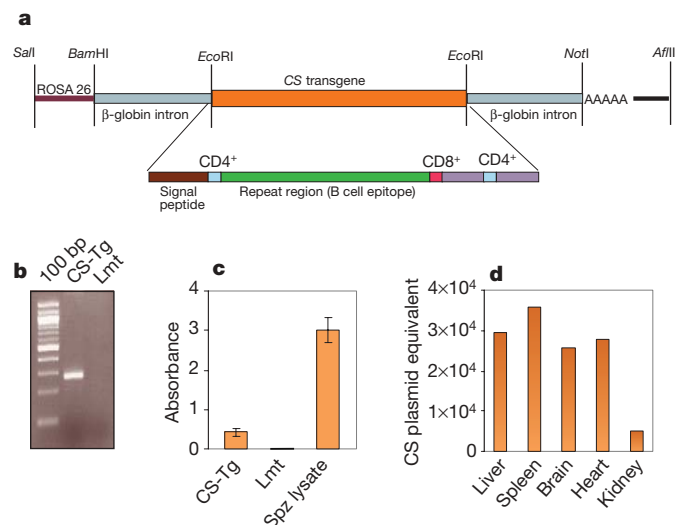


Figure 1 | Expression of CS in transgenic mice. **a**, Representation of the CS transgene depicting functional elements for mammalian expression, the localization of the T-cell CD4⁺ and CD8⁺ epitopes, and the B-cell immunodominant epitope. **b**, PCR showing the 277 bp transgene product amplified from gDNA of transgenic mice but not from non-transgenic littermates (Lmt). **c**, CS expression in the blood detected by ELISA that recognizes the repeats of CS. Sporozoite lysate (Spz lysate) is a positive control. Results are expressed as mean $\pm$ s.d. of absorbance values (at 405 nm) obtained from five readings per group. **d**, Tissue-specific levels of transgene mRNA as detected by qRT-PCR.

¹Michael Heidelberger Division of Immunology, Department of Pathology, New York University School of Medicine, New York, New York 10016, USA. ²Department of Microbiology and Immunology, School of Medicine, Keio University, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan. ³Laboratory of Molecular Immunology and Howard Hughes Institute, Rockefeller University, New York, New York 10021, USA. ⁴Department of Medical and Molecular Parasitology, and Department of Pathology, New York University School of Medicine, New York, New York 10016, USA. ⁵Department of Molecular Microbiology and Immunology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, USA.

*These authors contributed equally to this work.

the antibody response was predominantly IgM, whereas in the littermates and control BALB/c mice it was IgG (Fig. 4a). In the absence of CS-helper-T-cells, the T-cell-independent IgM response in the CS transgenic (CS-Tg) mice was consistent with the presence of a CS protein coat on the sporozoite surface that contains extended repeat domains. To demonstrate that the lack of class switching was indeed due to the absence of T-cell help, we transferred purified B cells (Supplementary Fig. 1) from the spleens of immunized CS transgenics and controls into JhT mice that have normal T cells, but are unable to make antibodies. As seen in Fig. 4b, a boost with *P. yoelii* IrSp induced production of IgG antibodies against the CS repeats by the transgenic B cells. Therefore transgenic expression of CS is not sufficient to delete or edit anti-CS specific B cells, and these cells produce T-cell-independent antibodies in response to immunization with IrSp.

Antibodies or IFN- γ -producing T cells specific for CS can protect against malaria infection¹⁴. To determine whether the T-cell-independent antibodies produced by the CS transgenics were biologically active, we tested them for the ability to induce shedding of CS from the surface of live sporozoites (CSP reaction), and to protect mice by passive transfer. The pooled serum of immunized CS transgenics elicited a strong CSP reaction (Supplementary Fig. 2), and blocked infection of naive mice with sporozoites (Fig. 4c). Western analysis of sporozoite extracts developed with immune serum from CS transgenic mice revealed only the two characteristic CS bands (Supplementary Fig. 3).

CS transgenics, control littermates and BALB/c mice were protected from challenge with intravenous injection of 2×10^4 sporozoites when immunized and boosted with 2×10^4 IrSp (Fig. 5a). In this paper we define 'protection' as the inhibition of development of EEFs as determined by parasite 18S rRNA copy number by quantitative PCR with reverse transcription (qRT-PCR). To examine the role of T cells in protection against non-CS antigens, we depleted CD8⁺ and CD4⁺ T cells from immunized CS transgenic mice before challenge. Figure 5b shows that the overall effect of depletion was small. Removal of CD8⁺ cells was, however, significantly more effective

than that of CD4⁺ cells. This experiment indicates that in the immunized CS transgenic mice there are only small numbers of non-CS, protective CD8⁺ cells.

To determine whether CS-specific T cells are uniquely protective after immunization with IrSp, we produced B-cell-deficient CS transgenic mice by crossing them with JhT mice (CS-Tg JhT). The CS-Tg JhT, control JhT and normal BALB/c mice were immunized and boosted with either 2×10^4 or 1×10^5 *P. yoelii* IrSp and then challenged with 2×10^4 sporozoites after either 10 or 30 days. Whereas immunized BALB/c and JhT mice showed parasite burdens that were 10^7 - and 10^4 -fold lower than controls, respectively, protection in the CS-Tg JhT mice was reverted by more than 90% (Fig. 5c, d and e). In these and all subsequent experiments we included controls with the same backgrounds as those of immunized mice. We conclude that the effector T cells directed against CS account for 90% or more of the reduction in liver stage burden in BALB/c mice following immunization with *P. yoelii* IrSp.

Next, we tried to reveal the protective efficacy of additional non-CS sporozoite antigens by priming the CS-Tg JhT mice with 5×10^4 IrSp, followed by two identical booster doses separated by 12 days. The mice were challenged with 2×10^4 sporozoites 10 days after the last immunizing dose. This led to sterile immunity as documented by the absence of blood stage infections for 8 successive days. However, if the mice were pre-treated with two doses of a monoclonal antibody to CD8⁺ T cells at days -1 and 0, blood infections were detected 3 days post-challenge (Fig. 5h). Thus, the powerful non-CS protective antigen(s) belong predominantly to the CD8⁺ subset.

To determine whether the dominant role of CS in T-cell-mediated immunity was strain specific, we produced CS transgenics in the C57BL/6 background. H2-b mice do not recognize CTL epitopes in CS molecules and are notoriously more difficult to protect than BALB/c mice¹⁵. The comparison of Fig. 5a and 5f, indeed shows that an identical immunization protocol that elicited very high levels of protection in BALB/c mice was much less effective in C57BL/6 mice. To exclude the participation of antibodies in protection, we produced B-cell-deficient C57BL/6 CS transgenic mice by crossing them with mice that carry a targeted deletion in the μ locus (CS-Tg μ MT). The CS-Tg μ MT and controls were immunized and boosted with 10^5 IrSp and challenged with 2×10^4 live sporozoites as above. Whereas control mice showed a 60-fold reduction in parasite EEF burden after immunization, we found only a 3–4-fold drop in the CS-Tg μ MT mice (Fig. 5g). We conclude that CS has an important role in protection mediated by IrSp immunization irrespective of mouse strain differences. Complete protection against sporozoite challenge has also been obtained by immunization of rodents with attenuated sporozoites^{16–18}. Whether or not CS is also immunodominant in these models remains to be investigated.

The results obtained by IrSp immunization of BALB/c shows that effector CD8⁺ T cells can play a central part in protection against

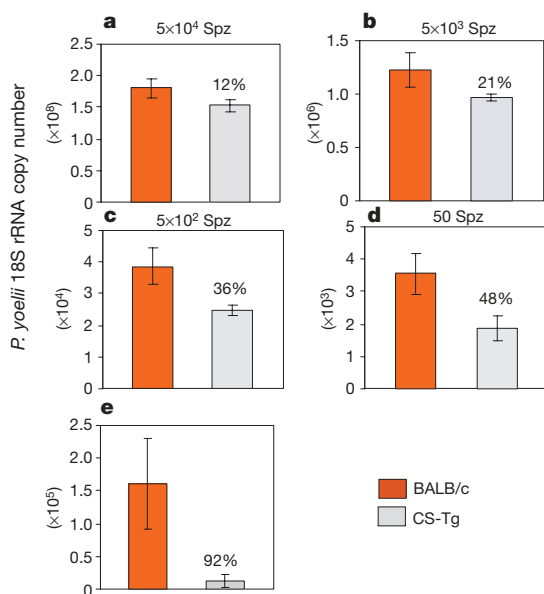


Figure 2 | Circulating CS in the transgenic mice inhibits infectivity of sporozoites. a–d, Groups of control BALB/c and CS-Tg mice were injected with varying doses of *P. yoelii* sporozoites (Spz). Percentage inhibition of infectivity was obtained by comparing the liver stage burdens at 40–42 h. e, Groups of BALB/c and CS-Tg mice were subjected to the bite of 10 mosquitoes for 5–6 min. Liver stage burden was evaluated by qRT-PCR of *P. yoelii* 18S rRNA copy number in all experiments. Results are expressed as mean $\pm$ s.d. of copy numbers from 4–5 mice per group.

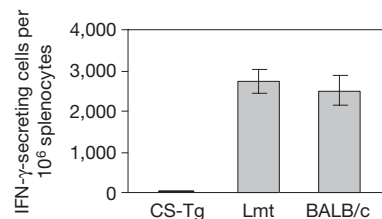


Figure 3 | CS-Tg mice are tolerant to the H-2K^d restricted CTL epitope of *P. yoelii* CS. CS-Tg, littermates (Lmt) and BALB/c mice were primed with 5×10^5 plaque forming units per mouse of recombinant yellow fever 17D vaccine containing the CD8⁺ T-cell epitope of *P. yoelii* CS¹³, and boosted with 1×10^7 pfu per mouse of modified vaccinia Ankara virus expressing the entire coding sequence of CS. The frequency of epitope-specific CD8⁺ T cells in spleens was measured by IFN- γ ELISPOT. Results are expressed as mean $\pm$ s.d. of epitope-specific CD8⁺ T cells from 4–5 mice per group.

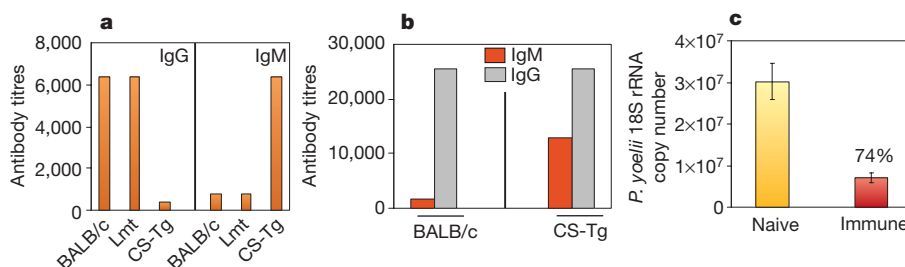


Figure 4 | CS-Tg mice produce T-cell-independent antibodies to CS. **a**, CS-Tg make predominantly antibodies to CS of IgM isotype. Serum samples from groups of sporozoite-immunized mice were analysed by ELISA against the recombinant *P. yoelii* CS repeat domain using peroxide labelled antibodies to isotype specific IgG or IgM. The titres are the highest dilution of serum having a change in absorbance (ΔA), at 405 nm, greater than the mean + 2 s.d. of non-immune sera. Lmt, CS non-transgenic littermate. **b**, B cells of CS-Tg mice make IgG antibodies with T-cell help. Purified B cells (2.5×10^7) from sporozoite-immunized BALB/c or CS-Tg mice were injected intravenously into JhT mice. Twenty-four hours later, the mice were

boosted with 2×10^4 IrSp. On day 10 post-immunization, sera were assayed for IgM and IgG levels against the CS repeat domain by ELISA. **c**, Passive transfer of CS-specific IgM from sporozoite immunized transgenic mice into normal BALB/c mice protect them from challenge. Two-hundred microlitres of immune sera from CS-Tg mice were injected intravenously into mice that were immediately challenged with 2×10^4 sporozoites. Protection was evaluated by comparison of *P. yoelii* 18S rRNA copy numbers from the livers of naive and recipient groups. Results are expressed as mean $\pm$ s.d. of copy numbers from 4–5 mice per group.

pre-erythrocytic forms of malaria. However, CD8⁺-T-cell-mediated protection is more difficult to achieve in malaria than in viral infections. One of the reasons is that EEFs are present exclusively in the liver and for only a few days. In contrast to viruses, which live in the cytoplasm of the host cells, the EEFs develop inside a parasitophorous vacuole. Thus, to provide the required targets for CTL recognition, EEF antigens must first cross the parasitophorous vacuole to reach the cytoplasm of hepatocytes. Recent studies have identified in *Plasmodium falciparum* a Pexel/VTS motif (*Plasmodium* export element, vacuolar transport signal) that is required for the transport of proteins across the parasitophorous vacuolar membrane^{19,20}. In addition to a signal sequence, the CS from all *Plasmodium* species contains one or two functional Pexel/VTS motifs (A. P. Singh and V. Nussenzweig,

unpublished data). These studies provide an explanation of the ability of CS to cross the parasitophorous vacuolar membrane.

Nevertheless, there are alternative mechanisms to generate 'protective' T-cell-epitopes. During sporozoite migration²¹ CS and TRAP/SSP2 (refs 22, 23) are left in trails in the cytoplasm of the traversed cells²⁴. It is conceivable that T cells that react with the CD4⁺ and CD8⁺ epitopes on Kupffer cells²⁵ or on hepatocytes that are close to the developing EEFs will secrete IFN- γ and inhibit parasite development (Supplementary Figs 4 and 5).

CS-based vaccines are already undergoing human trials. The RTS,S vaccine contains *P. falciparum* CS and has been extensively used in

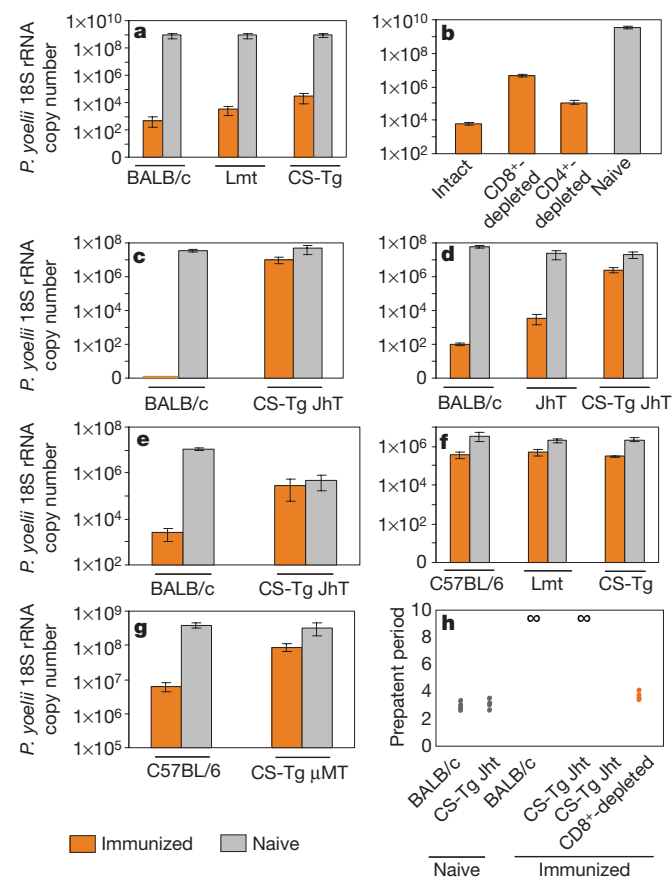


Figure 5 | Protection mediated by immunization with IrSp in BALB/c and C57BL/6 CS-Tg mice. **a–g**, In all experiments, protection was evaluated by qRT-PCR of *P. yoelii* 18S rRNA copy number in the liver following challenge with 2×10^4 normal sporozoites. Results are expressed as mean $\pm$ s.d. of copy numbers from 4–5 mice per group. **a**, Groups of control, CS-Tg and littermates in a BALB/c background were immunized and boosted with 2×10^4 *P. yoelii* IrSp. After challenge all groups were protected. Protection in CS-Tg mice was obtained in the absence of CS-specific T-cell help and CTL. Lmt, CS non-transgenic littermates. **b**, Groups of BALB/c CS-Tg mice were immunized and challenged as above. In two groups, either CD8⁺ or CD4⁺ T cells were depleted before challenge. Protection in CS-Tg mice was reverted only partially by T-cell depletion. **c**, BALB/c and CS-Tg JhT mice were immunized and boosted with higher doses (10^5) of *P. yoelii* IrSp and challenged as above. The control BALB/c mice were completely protected, whereas protection in CS-Tg JhT mice was reverted by 90%. **d**, The BALB/c, JhT and CS-Tg JhT mice were immunized and boosted with 2×10^4 *P. yoelii* IrSp and challenged as above. Protection was reverted in CS-Tg JhT mice. Some degree of protection was also observed in the absence of antibodies (JhT). **e**, BALB/c and CS-Tg JhT mice were immunized and boosted with 10^5 IrSp and challenged after 4 weeks. Liver burden in the CS-Tg JhT mice was almost identical to that of the control non-immunized mice, whereas immunized BALB/c mice were protected. **f**, Groups of control, CS-Tg and littermates in a C57BL/6 background were immunized and boosted with 2×10^4 *P. yoelii* IrSp. After challenge all groups of immunized animals had liver stage burden 8–9-fold lower as compared with naive groups. Nevertheless, the degree of protection was considerably smaller than that obtained in BALB/c mice (**a**). Lmt, CS non-transgenic littermates. **g**, C57BL/6 and CS-Tg μ MT mice were immunized and boosted with a higher dose (10^5) of *P. yoelii* IrSp and challenged as above. In contrast to BALB/c (**c**), the protection obtained in C57BL/6 was only partial, but it was reverted in CS-Tg μ MT mice. **h**, BALB/c and CS-Tg JhT mice were immunized and boosted twice with 5×10^4 IrSp. Some immunized CS-Tg JhT mice were depleted of CD8⁺ T cells before challenge. All groups were monitored for blood-stage infections from day 2 to 8 post-challenge. Whereas naive BALB/c, CS-Tg JhT and immunized CS-Tg JhT depleted of CD8⁺ T cells became positive for infection of blood stages between days 2 and 3 post-challenge, blood infections were not detected in immunized BALB/c and CS-Tg JhT for 8 successive days (∞).

naive individuals and in areas of Africa where malaria is endemic. In the latest published trial, about 30% of children vaccinated with RTS,S were protected against infection, and 50% against severe disease²⁶. Although the mechanism of protection is still being investigated, it appears that both antibodies against the CS repeats, and CD4⁺ T cells against a universal epitope of CS, participate in the RTS,S-mediated protection²⁷.

Other important findings are that sterile immunity can be obtained in the CS-Tg JhT BALB/c mice hyperimmunized with IrSp, and that protection is mediated by CD8⁺ T cells. The availability of CS transgenic mice should facilitate the identification of the non-CS protective antigen(s) in the rodent model. The human orthologue(s) may be additional malaria vaccine candidates.

METHODS

CS-transgene construction. For efficient expression of the *P. yoelii* 17X NL CS gene (NCBI accession number J02695) in mice, we designed an artificial *P. yoelii* CS gene that encodes amino acids with mouse codon usage²⁸. The artificial CS gene was designed with two point mutations at amino acids 25 and 328 (Asn→Gln) to avoid glycosylation in the mouse. The sequence encoding the carboxy-terminal region corresponding to the glycosyl phosphatidylinositol attachment sequence (from amino acid 347) was removed from the gene. After the artificial CS gene was sequenced and confirmed, it was inserted into the *EcoRI* site of the rabbit β -globin intron polyA DNA fragment (NCBI accession number K03256, from 903 bp to 2063 bp). Restriction sites *Bam*HI and *Not*I were introduced by PCR at the 5' and 3' ends of the transgene construct. Plasmid pEGFP-N1 (Clontech) was used as a backbone for ligating the transgene using *Bam*HI and *Not*I sites, and replacing the original *EGFP* gene present in the multiple cloning site of the plasmid. ROSA26 promoter sequence was introduced upstream of the CS transgene using *Sal*I and *Bam*HI sites. The plasmid sequence was removed by *Afl*II and *Sal*I restriction digestion. The transgenes were co-injected into C57BL/6 oocytes and the resulting transgenic mice were identified by PCR screening of genomic DNA obtained from tails.

Screening of CS-Tg mice. Screening of CS-Tg mice was done by PCR of gDNA obtained from tail samples using primers CS-TgF 5'-GGCGCAGAGAAACCTCAACGAACTATGCTATAACG-3' and CS-TgR 5'-CTTGGGGGGTCTGCTCC TTCTTCGGAT-3'. The PCR detected a 277 bp product specific to all transgenic mice.

Expression of CS in transgenic mice. CS expression in transgenic mice was demonstrated by ELISA and qRT-PCR methods as described in Supplementary Information.

Immunization and challenge. Details of immunization regimen and challenge doses used for different experiments are described in Supplementary Information.

Detection and quantification of *P. yoelii* 18S rRNA copy number by qRT-PCR. Quantification of *P. yoelii* 18S rRNA copy number was performed essentially as described earlier²⁹ except that a double-stranded-DNA-specific iQ SYBR Green super mix (Biorad laboratories) was used to detect the PCR product. Copy numbers were calculated using known number of *P. yoelii* 18S rRNA plasmid standard.

Quantification of epitope-specific CD8⁺ T cells by ELISPOT assay. ELISPOT assay was performed as previously described³⁰. For these assays we used MHC-compatible A20.2J target cells coated with SYVPSAEQI peptide that react with CS-specific CD8⁺ T cells.

T-cell depletion experiments. Monoclonal antibodies to CD4⁺ (GK1.5) and CD8⁺ (YTS169) were used to deplete T-cell subsets from immunized CS-Tg mice. Depletion was performed by intraperitoneal administration of 150 μ g of respective antibodies for 4 consecutive days before challenge with viable sporozoites.

GST recombinant protein for *P. yoelii* CS repeats, serologic assays, passive transfer of immune sera, passive transfer of B cells and antibody-deficient mice. Complete details are given in Supplementary Information.

Received 17 August; accepted 20 October 2006.

Published online 6 December 2006.

1. Nussenzweig, R. S., Vanderberg, J., Most, H. & Orton, C. Protective immunity produced by the injection of x-irradiated sporozoites of *Plasmodium berghei*. *Nature* **216**, 160–162 (1967).
2. Gwadz, R. W., Cochrane, A. H., Nussenzweig, V. & Nussenzweig, R. S. Preliminary studies on vaccination of rhesus monkeys with irradiated sporozoites of *Plasmodium knowlesi* and characterization of surface antigens of these parasites. *Bull. World Health Organ.* **57** (Suppl 1), 165–173 (1979).
3. Clyde, D. F., Most, H., McCarthy, V. C. & Vanderberg, J. P. Immunization of man against sporozoite-induced *falciparum* malaria. *Am. J. Med. Sci.* **266**, 169–177 (1973).

4. Romero, P. *et al.* Cloned cytotoxic T cells recognize an epitope in the circumsporozoite protein and protect against malaria. *Nature* **341**, 323–326 (1989).
5. Kappe, S. H. *et al.* Exploring the transcriptome of the malaria sporozoite stage. *Proc. Natl Acad. Sci. USA* **98**, 9895–9900 (2001).
6. Wang, Q., Brown, S., Roos, D. S., Nussenzweig, V. & Bhanot, P. Transcriptome of asexual liver stages of *Plasmodium yoelii*. *Mol. Biochem. Parasitol.* **137**, 161–168 (2004).
7. Hoffman, S. L. *et al.* Protection of humans against malaria by immunization with radiation-attenuated *Plasmodium falciparum* sporozoites. *J. Infect. Dis.* **185**, 1155–1164 (2002).
8. Hoffman, S. L. & Doolan, D. L. Malaria vaccines—targeting infected hepatocytes. *Nature Med.* **6**, 1218–1219 (2000).
9. Hoffman, S. L., Rogers, W. O., Carucci, D. J. & Venter, J. C. From genomics to vaccines: malaria as a model system. *Nature Med.* **4**, 1351–1353 (1998).
10. Frevert, U. *et al.* Malaria circumsporozoite protein binds to heparan sulfate proteoglycans associated with the surface membrane of hepatocytes. *J. Exp. Med.* **177**, 1287–1298 (1993).
11. Sinnis, P. *et al.* Structural and functional properties of region II-plus of the malaria circumsporozoite protein. *J. Exp. Med.* **180**, 297–306 (1994).
12. Wang, Q., Fujioka, H. & Nussenzweig, V. Exit of *Plasmodium* sporozoites from oocysts is an active process that involves the circumsporozoite protein. *PLoS Pathog.* **1**, e9 (2005).
13. Tao, D. *et al.* Yellow fever 17D as a vaccine vector for microbial CTL epitopes: protection in a rodent malaria model. *J. Exp. Med.* **201**, 201–209 (2005).
14. Schofield, L. *et al.* γ Interferon, CD8⁺ T cells and antibodies required for immunity to malaria sporozoites. *Nature* **330**, 664–666 (1987).
15. Weiss, W. R., Good, M. F., Hollingdale, M. R., Miller, L. H. & Berzofsky, J. A. Genetic control of immunity to *Plasmodium yoelii* sporozoites. *J. Immunol.* **143**, 4263–4266 (1989).
16. Mueller, A. K., Labaied, M., Kappe, S. H. & Matuschewski, K. Genetically modified *Plasmodium* parasites as a protective experimental malaria vaccine. *Nature* **433**, 164–167 (2005).
17. Mueller, A. K. *et al.* *Plasmodium* liver stage developmental arrest by depletion of a protein at the parasite–host interface. *Proc. Natl Acad. Sci. USA* **102**, 3022–3027 (2005).
18. van Dijk, M. R. *et al.* Genetically attenuated, P36p-deficient malarial sporozoites induce protective immunity and apoptosis of infected liver cells. *Proc. Natl Acad. Sci. USA* **102**, 12194–12199 (2005).
19. Marti, M., Good, R. T., Rug, M., Knuepfer, E. & Cowman, A. F. Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* **306**, 1930–1933 (2004).
20. Hiller, N. L. *et al.* A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* **306**, 1934–1937 (2004).
21. Mota, M. M. *et al.* Migration of *Plasmodium* sporozoites through cells before infection. *Science* **291**, 141–144 (2001).
22. Wikel, B. *et al.* Irradiated sporozoite vaccine induces HLA-B8-restricted cytotoxic T lymphocyte responses against two overlapping epitopes of the *Plasmodium falciparum* sporozoite surface protein 2. *J. Exp. Med.* **182**, 1435–1445 (1995).
23. Robson, K. J. *et al.* A highly conserved amino-acid sequence in thrombospondin, properdin and in proteins from sporozoites and blood stages of a human malaria parasite. *Nature* **335**, 79–82 (1988).
24. Sultan, A. A. *et al.* TRAP is necessary for gliding motility and infectivity of plasmodium sporozoites. *Cell* **90**, 511–522 (1997).
25. Frevert, U. *et al.* Intravital observation of *Plasmodium berghei* sporozoite infection of the liver. *PLoS Biol.* **3**, e192 (2005).
26. Alonso, P. L. *et al.* Duration of protection with RTS, S/AS02A malaria vaccine in prevention of *Plasmodium falciparum* disease in Mozambican children: single-blind extended follow-up of a randomised controlled trial. *Lancet* **366**, 2012–2018 (2005).
27. Reece, W. H. *et al.* A CD4⁺ T-cell immune response to a conserved epitope in the circumsporozoite protein correlates with protection from natural *Plasmodium falciparum* infection and disease. *Nature Med.* **10**, 406–410 (2004).
28. Sano, G., Morimatsu, K. & Horii, T. Purification and characterization of dihydrofolate reductase of *Plasmodium falciparum* expressed by a synthetic gene in *Escherichia coli*. *Mol. Biochem. Parasitol.* **63**, 265–273 (1994).
29. Bruna-Romero, O. *et al.* Detection of malaria liver-stages in mice infected through the bite of a single *Anopheles* mosquito using a highly sensitive real-time PCR. *Int. J. Parasitol.* **31**, 1499–1502 (2001).
30. Carvalho, L. H., Hafalla, J. C. & Zavala, F. ELISPOT assay to measure antigen-specific murine CD8⁺ T cell responses. *J. Immunol. Methods* **252**, 207–218 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements V.N. is supported by the Gates Foundation, R.S.N. by the Starr Foundation, F.Z. by the NIH, and M.C.N. is a Howard Hughes Investigator. We thank A. Nussenzweig for suggestions. S.B. is supported by the Pew Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.K. (arun.kumar@med.nyu.edu).

Inhibition of cytohesins by SecinH3 leads to hepatic insulin resistance

Markus Hafner¹, Anton Schmitz¹, Imke Grüne¹, Seergazhi G. Srivatsan¹, Bianca Paul², Waldemar Kolanus², Thomas Quast², Elisabeth Kremmer³, Inga Bauer¹ & Michael Famulok¹

G proteins are an important class of regulatory switches in all living systems. They are activated by guanine nucleotide exchange factors (GEFs), which facilitate the exchange of GDP for GTP^{1,2}. This activity makes GEFs attractive targets for modulating disease-relevant G-protein-controlled signalling networks^{3–5}. GEF inhibitors are therefore of interest as tools for elucidating the function of these proteins and for therapeutic intervention; however, only one small molecule GEF inhibitor, brefeldin A (BFA), is currently available^{6–9}. Here we used an aptamer displacement screen to identify SecinH3, a small molecule antagonist of cytohesins. The cytohesins are a class of BFA-resistant small GEFs for ADP-ribosylation factors (ARFs), which regulate cytoskeletal organization¹⁰, integrin activation¹¹ or integrin signalling¹². The application of SecinH3 in human liver cells showed that insulin-receptor-complex-associated cytohesins are required for insulin signalling. SecinH3-treated mice show increased expression of gluconeogenic genes, reduced expression of glycolytic, fatty acid and ketone body metabolism genes in the liver, reduced liver glycogen stores, and a compensatory increase in plasma insulin. Thus, cytohesin inhibition results in hepatic insulin resistance. Because insulin resistance is among the earliest pathological changes in type 2 diabetes, our results show the potential of chemical biology for dissecting the molecular pathogenesis of this disease.

Cytohesins are multi-domain proteins in which a Sec7 domain bears the GEF activity. Unlike the large (~200 kDa) ARF-GEFs, the small ones (~47 kDa) are insensitive to BFA^{13,14}. The only known small GEF Sec7-domain inhibitor is the RNA aptamer M69, which represses guanine nucleotide exchange *in vitro* and reduces T-cell adhesion¹⁰. Using M69 and the cytohesin-1 Sec7 domain, we established an assay based on fluorescence polarization to identify cytohesin-specific small molecules that displace the aptamer from its target and adopt its inhibitory activity (see Supplementary Fig. 1a). From a diverse library of synthetic chemicals, we identified a series of 1,2,4-triazole derivatives as initial hits and selected the most promising compound, H3 (Fig. 1a, left), for synthesis and further studies (see Supplementary Figs 1, 2). H3 bound to the Sec7 domains of human cytohesins-1–3 with dissociation constants (K_d values) between 200 and 250 nM (Table 1). H3e, H3ip and H3bio, in which substituent R_1 is increased (see Supplementary Fig. 2), showed gradually decreased Sec7 affinity. The Sec7 domain of EFA6, an ARF6-GEF that does not belong to the cytohesin family, was bound about 30-fold more weakly, despite its 32% identity and 46% similarity with the cytohesin-1 Sec7 domain. The CDC25-like domain of the ras-GEF Son of sevenless (Sos) was not bound.

To assess the compound's inhibitory potential, guanine nucleotide exchange assays were performed with ARF1 and full-length

cytohesins^{15,16}. All cytohesins, including the mouse and *Drosophila* homologues of cytohesin-3, exhibited half-maximal inhibitory concentrations (IC_{50} values) between 2.4 and 5.6 μ M (Table 1), indicating that H3 recognizes an evolutionarily conserved region. For EFA6–Sec7 and Gea2–Sec7, a large GEF from yeast, the IC_{50} s were 18- and 12-fold higher, respectively. The GEF activity of Sos was unaffected. Because of its inhibitory profile, we named the new compound class Secins (for Sec7 inhibitors).

To test whether SecinH3 maintained its *in vitro* preference for cytohesins in living cells, we monitored its effect on the structural integrity of the Golgi apparatus, which depends on the function of BFA-sensitive large GEFs¹⁷. Treatment with 20 μ M BFA completely disrupted Golgi integrity (Fig. 1b, panel 2), whereas treatment with SecinH3 caused a minor effect only at or beyond 50 μ M, consistent with this concentration being close to the IC_{50} for the large GEF Gea2 (Fig. 1b, panels 3, 4). In summary, these data show that SecinH3 is a Sec7-specific GEF inhibitor with preference for the small GEFs of the cytohesin family.

Four different but highly homologous cytohesins are known in mammals, whereas only one, the cytohesin homologue steppke,

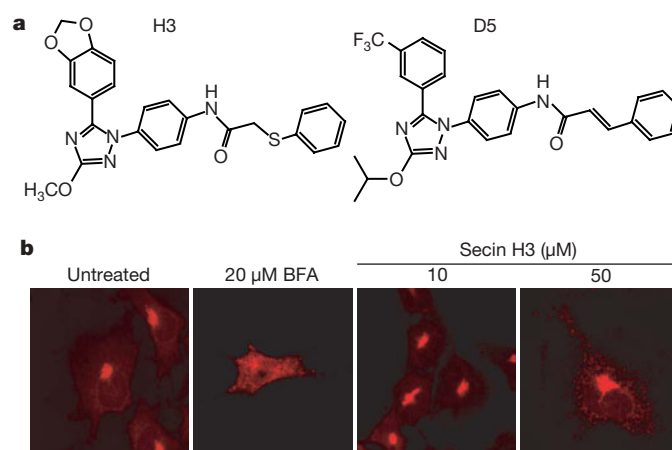


Figure 1 | Structure and characterization of SecinH3. **a**, Molecular structures of SecinH3 (left) and the negative control compound D5 (right). For details on the synthesis and characterization of each compound, see Supplementary Methods and Supplementary Fig. 2. **b**, SecinH3 does not affect Golgi structure. Compared to untreated cells, no significant disturbance of Golgi integrity is observed at SecinH3 concentrations of 10 μ M and 50 μ M, respectively, whereas BFA leads to complete Golgi destruction at 20 μ M. Golgi membranes were visualized using an anti-galactin antibody.

¹LIMES Program Unit Chemical Biology & Medicinal Chemistry, c/o Kekulé Institut für Organische Chemie und Biochemie, University of Bonn, Gerhard-Domagk-Straße 1, 53121 Bonn, Germany. ²LIMES Program Unit Molecular Immune and Cell Biology, Laboratory of Molecular Immunology, University of Bonn, Karlrobert-Kreiten Straße 13, 53115 Bonn, Germany. ³Institut für Molekulare Immunologie, GSF-Forschungszentrum für Umwelt und Gesundheit, Marchioninistr. 25, 81377 München, Germany.

Table 1 | SecinH3 binds cytohesin Sec7 domains and inhibits GDP/GTP exchange

Protein	Dissociation constant, K_d (nM)*				Inhibition of GDP/GTP exchange, IC_{50} (μ M)†		
	H3	H3e	H3ip	H3bio	Protein	H3	D5
hCyh1-S7	200 $\pm$ 10				hCyh1	5.4 $\pm$ 0.3	n.d.‡
hCyh2-S7	250 $\pm$ 5	305 $\pm$ 15	370 $\pm$ 18	523 $\pm$ 26	hCyh2	2.4 $\pm$ 0.1	n.d.‡
hCyh3-S7	239 $\pm$ 17				hCyh3	5.6 $\pm$ 0.4	
					mCyh3	5.4 $\pm$ 0.6	
					steppke§	5.6 $\pm$ 0.2	
					yGea2-S7	65 $\pm$ 4.0	
hEFA6-S7	7,600 $\pm$ 200				hEFA6-S7	>100	
hSos	n.d.‡				hSos	n.d.‡	

* Measured by isothermal titration calorimetry (see Supplementary Fig. 3).

† GTP exchange of the indicated full-length GEFs or GEF-Sec7 domains on ARF1 was measured using the tryptophan fluorescence assay.

‡ n.d., not detectable.

§ *Drosophila* homologue of cytohesins.

|| Sos-catalysed GTP exchange on Ras was measured by the mantGDP assay.

exists in flies. So far, no cytohesin knockout mouse has been described, but flies lacking *steppke* show reduced growth¹⁸, suggesting that it is involved in insulin signalling. We tested the requirement for mammalian cytohesins in insulin signalling by monitoring the transcription of the prototypic insulin-regulated gene for insulin-like growth factor binding protein 1 (*IGFBP1*) in the human liver cell line HepG2. SecinH3 almost completely blocked the insulin-dependent transcriptional repression of *IGFBP1* (Fig. 2a) with an IC_{50} of 2.2 μ M (see Supplementary Fig. 4; see Supplementary Information and

Nature Protocols doi: 10.1038/nprot.2006.413 for details of methods). D5, a structurally related compound (Fig. 1a, right) that was inactive in the aptamer displacement and GDP/GTP exchange assays (see Supplementary Fig. 1b; Table 1) did not affect insulin-dependent levels of *IGFBP1* mRNA (Fig. 2a). To analyse whether SecinH3 affects ARF activation by cytohesins in living cells we used immobilized GGA3 (Golgi-associated, gamma adaptin ear containing, ARF binding protein 3) to capture activated, GTP-bound ARF¹⁹ (see Supplementary Methods and *Nature Protocols* doi: 10.1038/nprot.2006.412 for detailed methods). Insulin stimulation increased the level of ARF6-GTP that was pulled down with GGA3, which was reduced in SecinH3-treated cells but remained unaffected by D5 (Fig. 2b). Consistently, insulin-stimulated translocation of ARF6 to the plasma membrane was also inhibited by SecinH3 (see Supplementary Fig. 5). Finally, SecinH3 had no direct effect on various kinases, including those involved in insulin signalling (Supplementary Table 1).

To confirm that SecinH3 affects insulin signalling by inhibition of cytohesins, we knocked down cytohesins 1–3 (*cyh1*–*3*), the only expressed cytohesins in HepG2 cells (data not shown), by RNA interference and analysed gene expression (Fig. 2c). Quantitative PCR (qPCR) and western blotting confirmed the specificity of siRNAs for their respective target mRNAs (see Supplementary Fig. 6). Whereas the knock-down of *cyh1* did not prevent the insulin-dependent repression of *IGFBP1* expression, knockdown of *cyh3* inhibited the insulin effect to a similar level as with SecinH3, and knockdown of *cyh2* had an even stronger effect. An ARF6-specific siRNA also inhibited the insulin effect whereas an ARF1-specific siRNA did not (see Supplementary Fig. 7).

Expression of *IGFBP1* is controlled by insulin acting through the forkhead box transcription factors FoxO1A and FoxO3A. Upon insulin stimulation, protein kinase B (PKB/Akt) is activated by phosphorylation and translocates to the nucleus, where it phosphorylates FoxO proteins^{20,21} leading to their nuclear exclusion and, thus, reduction of target gene expression^{22,23}. We found that SecinH3 inhibited the insulin-dependent phosphorylation of Akt and FoxO1A in a concentration-dependent manner (Fig. 2d; Supplementary Fig. 8). Morphological analysis of HepG2 cells transfected with enhanced green fluorescent protein (EGFP)-tagged FoxO1A showed that the insulin-induced exclusion of FoxO1A from the nucleus was completely prevented by SecinH3 (see Supplementary Fig. 9).

Metabolic insulin signalling in the target cell is initiated by insulin binding to the insulin receptor, which undergoes autophosphorylation on cytoplasmic tyrosine residues. This is followed by binding and phosphorylation of specific tyrosines in the insulin receptor substrate (IRS) proteins. The phosphotyrosines of pIRS serve as binding sites for phosphatidylinositol-3-OH kinase (PI(3)K), which then activates its downstream targets, including Akt²⁴. Neither autophosphorylation of the insulin receptor nor its density on the HepG2 cell surface was affected by SecinH3 (see Supplementary Fig. 10), showing that cytohesins act after receptor activation. However, the next step, activation of IRS1 by its phosphorylation, was inhibited

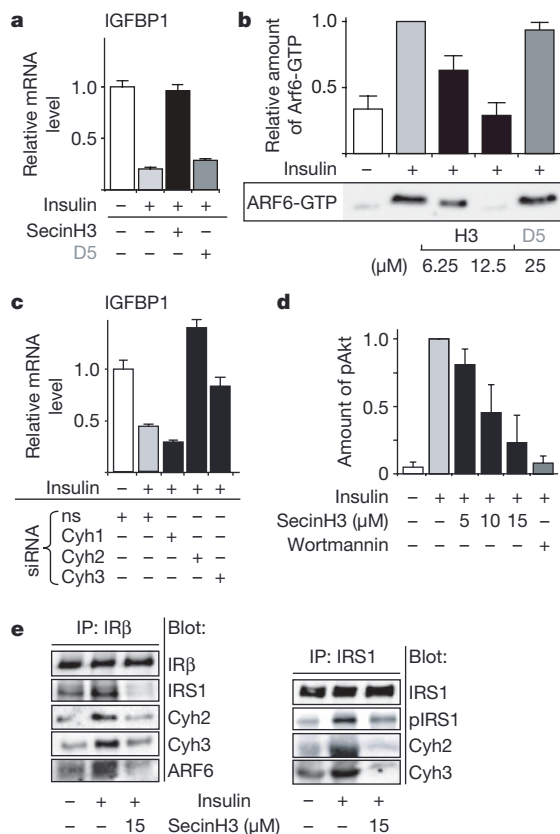


Figure 2 | SecinH3 inhibits insulin signalling in HepG2 cells. a, Effect of SecinH3 on insulin-dependent *IGFBP1* expression, determined by quantitative PCR (qPCR) ($n = 3$). **b**, Insulin-induced ARF activation is inhibited by SecinH3 ($n = 4$). Lower panel: representative western blot of activated ARF6. **c**, Cells were transfected with the indicated siRNA or non-silencing control siRNA (ns) and *IGFBP1* expression was determined by qPCR ($n = 3$). **d**, SecinH3 reduces Akt phosphorylation. pAkt was detected by immunoblot (see Supplementary Fig. 8), using actin for normalization ($n > 6$). **e**, Effect of SecinH3 on insulin receptor and IRS1 complex formation. IR β (left) and IRS1 (right) were immunoprecipitated, and coprecipitated proteins were detected by immunoblotting. Error bars: s.d.

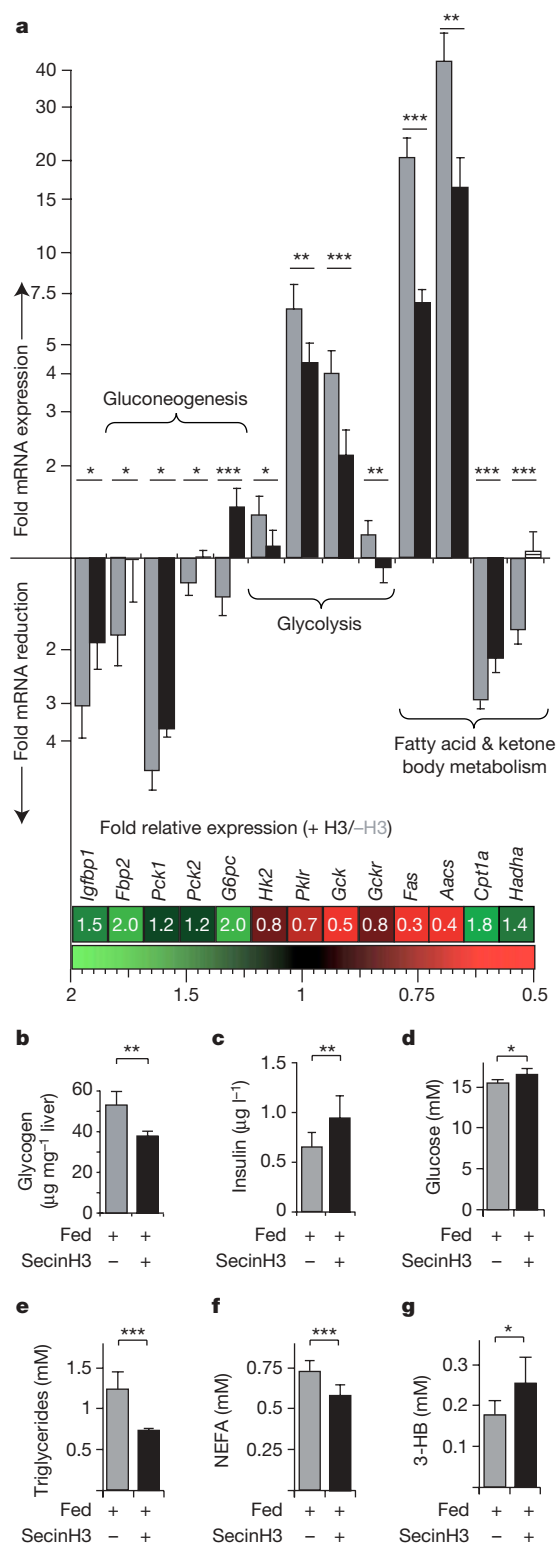


Figure 3 | Impaired cytohesin function results in hepatic insulin resistance.

a, Gene expression in liver was determined by qPCR in mice fed with (black) or without SecinH3 (grey). Upper part: change in gene expression as compared to starved animals (set as 1). Lower part: ratio of gene expression in SecinH3-fed versus control animals ($n = 6$). Insulin-repressed genes are overexpressed (green); insulin-induced genes are underexpressed (red). **b**, Glycogen levels³¹ in livers of SecinH3-fed and control mice, expressed as μg glucose units per mg liver ($n = 5$). Serum levels of insulin (**c**), glucose (**d**), triglycerides (**e**), non-esterified fatty acids (NEFA) (**f**) and 3-hydroxybutyrate (3-HB) (**g**) were determined in non-starved SecinH3-fed and control mice ($n = 8$). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Error bars: s.d.

by SecinH3 (see Supplementary Fig. 11). Co-immunoprecipitation studies revealed insulin-dependent interactions of *cyh2*, *cyh3*, and ARF6 with the insulin receptor–IRS1 complex, which were reduced in the presence of SecinH3 (Fig. 2e). The binding of IRS1 to the insulin receptor was also inhibited by SecinH3, indicating that the cytohesin–ARF6 complex facilitates the formation of the insulin receptor–IRS1 complex. These results show that *cyh2* and *cyh3* are fundamentally involved in the earliest steps of insulin signalling. We propose that the insulin-dependent physical association of the cytohesin–ARF6 complex with the insulin receptor–IRS complex is required for the phosphorylation of IRS and thus for the activation of the insulin signalling cascade immediately at the insulin receptor.

We next investigated the involvement of cytohesins in insulin signalling *in vivo*, by feeding mice chow containing SecinH3. Liquid chromatography–mass spectrometry analysis verified that livers from SecinH3-fed mice contained the unaltered compound (see Supplementary Fig. 12; method is described in detail in Supplementary Information and *Nature Protocols* doi: 10.1038/nprot.2006.414). Besides *Igf1*, we analysed hepatic expression of the genes for pyruvate carboxylase-1 and -2 (*Pck1*, *Pck2*), fructose-1,6-bisphosphatase 2 (*Fbp2*) and glucose-6-phosphatase (*G6pc*), which are involved in gluconeogenesis, and the glycolytic enzymes glucokinase (*Gck*), its regulator (*Gckr*), pyruvate kinase (*Pklr*), and hexokinase 2 (*Hk2*). Compared to mice fed the same chow without SecinH3, the expression levels of the insulin-repressed gluconeogenic genes were elevated (green in Fig. 3a), whereas the insulin-induced glycolytic genes were reduced (red in Fig. 3a). Insulin-stimulated Akt phosphorylation was also inhibited in SecinH3-treated mice (see Supplementary Fig. 13). Analysis of liver glycogen levels revealed that the compound also inhibited insulin-dependent glycogen synthesis (Fig. 3b). SecinH3 also inhibited the insulin-induced expression of the genes for fatty acid synthase (*Fasn*) and acetoacetyl-CoA synthetase (*Aacs*), which are involved in triglyceride synthesis and ketone body metabolism, respectively (Fig. 3a). The expression of the genes for two key enzymes of mitochondrial β -oxidation, carnitine palmitoyltransferase 1a (*Cpt1a*) and hydroxyacyl-CoA dehydrogenase (*Hadha*), both of which are repressed by insulin, was increased in the SecinH3-treated mice (Fig. 3a).

All available biochemical and gene expression data therefore converge on the view that cytohesins are fundamentally involved in insulin signalling and that their impairment results in physiological alterations that are characteristic of hepatic insulin resistance. Insulin resistance is an important hallmark of the pre-clinical stages of type 2 diabetes; organisms compensate for this by increasing insulin secretion to maintain blood glucose at physiological levels. We found significantly increased levels of serum insulin with slightly elevated glucose concentrations in SecinH3-treated mice (Fig. 3c, d). In accordance with the reduced *Fas* expression, serum triglycerides (Fig. 3e) and non-esterified fatty acids (Fig. 3f) were reduced. Enhanced β -oxidation due to elevated *Cpt1a* and *Hadha* levels, which leads to increased generation of acetyl-CoA together with reduced acetoacetate consumption by *Aacs*, should raise the levels of ketone bodies. Accordingly, 3-hydroxybutyrate was increased in the serum of SecinH3-treated mice (Fig. 3g). In summary, the effects of the cytohesin inhibitor SecinH3 on insulin signalling and gene expression in the liver, and the resulting physiological alterations, correspond to the pathological changes seen in mice with a liver-specific insulin receptor knockout²⁵ (see Supplementary Table 2), indicating that not only are cytohesins required for insulin signalling in the liver, but their dysfunction also contributes to the pathogenesis of hepatic insulin resistance.

Insulin resistance is considered to have a causative role in the pathogenesis of type 2 diabetes²⁶. However, the molecular mechanism that leads to impaired insulin sensitivity is poorly understood. Knockout of either the insulin receptor²⁵ or IRS-2 (refs 27, 28), or the liver-specific expression of a constitutively active form of FoxO1 (ref. 29) in transgenic mice, results in hepatic insulin resistance.

Because mutations of these genes have not been described for most patients suffering from type 2 diabetes³⁰, additional factors seem to be involved. Our data indicate that cytohesins might be one of these factors and that their function might be crucial for regulating the sensitivity of insulin-responsive tissues.

Furthermore, we show here how to translate information stored within an aptamer into a small molecule. Thereby, chemical space can be explored in a rapid, focused, and modular manner, by indirectly taking advantage of the highest molecular diversity currently amenable to screening, namely that of up to 10^{16} different nucleic acid sequences. Aptamer displacement screens demand small molecules of high initial potency, leading to effective and specific inhibitors that can be used as drug leads or tools in chemical biology.

METHODS

Details of aptamer labelling and the aptamer displacement screen, determination of K_{d} s by isothermal titration calorimetry, GDP/GTP exchange assays, the ARF6 activation assay, mice, quantification of *Igfbp1* expression in cell culture, quantification of gene expression in mouse liver, siRNA transfections, immunoblotting and immunoprecipitation, determination of physiological parameters, synthesis and characterization of compounds, and the characterization of the cytohesin-3 antibody are provided in Supplementary Information.

Statistics. Results are given as the mean $\pm$ s.d. Statistical analyses were performed by the two-tailed *t*-test using the InStat program (GraphPad). All data sets passed the Kolmogorov and Smirnov test for gaussian distribution. Where the standard deviations of the compared pairs were different, the Welch correction was applied to the *t*-test. Differences of means were considered significant at a significance level of 0.05.

Received 21 September; accepted 3 November 2006.

- Cherfils, J. & Chardin, P. GEFs: structural basis for their activation of small GTP-binding proteins. *Trends Biochem. Sci.* **24**, 306–311 (1999).
- Jackson, C. L. & Casanova, J. E. Turning on ARF: the Sec7 family of guanine-nucleotide-exchange factors. *Trends Cell Biol.* **10**, 60–67 (2000).
- Hashimoto, S. *et al.* Requirement for Arf6 in breast cancer invasive activities. *Proc. Natl Acad. Sci. USA* **101**, 6647–6652 (2004).
- Gonzalez-Garcia, A. *et al.* RalGDS is required for tumor formation in a model of skin carcinogenesis. *Cancer Cell* **7**, 219–226 (2005).
- Malliri, A. *et al.* Mice deficient in the Rac activator Tiam1 are resistant to Ras-induced skin tumours. *Nature* **417**, 867–871 (2002).
- Donaldson, J. G., Finazzi, D. & Klausner, R. D. Brefeldin A inhibits Golgi membrane-catalysed exchange of guanine nucleotide onto ARF protein. *Nature* **360**, 350–352 (1992).
- Helms, J. B. & Rothman, J. E. Inhibition by brefeldin A of a Golgi membrane enzyme that catalyses exchange of guanine nucleotide bound to ARF. *Nature* **360**, 352–354 (1992).
- Morinaga, N., Tsai, S.-C., Moss, J. & Vaughan, M. Isolation of a brefeldin A-inhibited guanine nucleotide-exchange protein for ADP ribosylation factor (ARF) 1 and ARF3 that contains a Sec7-like domain. *Proc. Natl Acad. Sci. USA* **93**, 12856–12860 (1996).
- Peyroche, A. *et al.* Brefeldin A acts to stabilize an abortive ARF–GDP–Sec7 domain protein complex: involvement of specific residues of the Sec7 domain. *Mol. Cell Biol.* **3**, 275–285 (1999).
- Mayer, G. *et al.* Controlling small guanine-nucleotide-exchange factor function through cytoplasmic RNA intramers. *Proc. Natl Acad. Sci. USA* **98**, 4961–4965 (2001).
- Kolanus, W. *et al.* $\alpha\text{L}\beta\text{2}$ integrin/LFA-1 binding to ICAM-1 induced by cytohesin-1, a cytoplasmic regulatory molecule. *Cell* **86**, 233–242 (1996).
- Perez, O. D. *et al.* Leukocyte functional antigen 1 lowers T cell activation thresholds and signaling through cytohesin-1 and Jun-activating binding protein 1. *Nature Immunol.* **4**, 1083–1092 (2003).
- Renault, L., Guibert, B. & Cherfils, J. Structural snapshots of the mechanism and inhibition of a guanine nucleotide exchange factor. *Nature* **426**, 525–530 (2003).
- Mossessova, E., Corpina, R. A. & Goldberg, J. Crystal structure of ARF1*Sec7 complexed with Brefeldin A and its implications for the guanine nucleotide exchange mechanism. *Mol. Cell* **12**, 1403–1411 (2003).
- Kruljac-Letic, A., Moelleken, J., Kallin, A., Wieland, F. & Blaukat, A. The tyrosine kinase Pyk2 regulates Arf1 activity by phosphorylation and inhibition of the Arf-GTPase-activating protein ASAP1. *J. Biol. Chem.* **278**, 29560–29570 (2003).
- Ahmadian, M. R., Wittinghofer, A. & Herrmann, C. Fluorescence methods in the study of small GTP-binding proteins. *Methods Mol. Biol.* **189**, 45–63 (2002).
- Moss, J. & Vaughan, M. Molecules in the ARF Orbit. *J. Biol. Chem.* **273**, 21431–21434 (1998).
- Fuss, B., Becker, T., Zinke, I. & Hoch, M. The cytohesin Steppke is essential for insulin signalling in *Drosophila*. *Nature* doi:10.1038/nature05412 (this issue).
- Takatsu, H., Yoshino, K., Toda, K. & Nakayama, K. GGA proteins associate with Golgi membranes through interaction between their GGAH domains and ADP-ribosylation factors. *Biochem. J.* **365**, 369–378 (2002).
- Brunet, A. *et al.* Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857–868 (1999).
- Kops, G. J. *et al.* Direct control of the Forkhead transcription factor AFX by protein kinase B. *Nature* **398**, 630–634 (1999).
- Burgering, B. M. & Kops, G. J. Cell cycle and death control: long live Forkheads. *Trends Biochem. Sci.* **27**, 352–360 (2002).
- Accili, D. & Arden, K. C. FoxOs at the crossroads of cellular metabolism, differentiation, and transformation. *Cell* **117**, 421–426 (2004).
- Knight, Z. A. *et al.* A pharmacological map of the PI3-K family defines a role for p110 α in insulin signaling. *Cell* **125**, 733–747 (2006).
- Michael, M. D. *et al.* Loss of insulin signaling in hepatocytes leads to severe insulin resistance and progressive hepatic dysfunction. *Mol. Cell* **6**, 87–97 (2000).
- Biddinger, S. B. & Kahn, C. R. From mice to men: insights into the insulin resistance syndromes. *Annu. Rev. Physiol.* **68**, 123–158 (2006).
- Kubota, N. *et al.* Disruption of insulin receptor substrate 2 causes type 2 diabetes because of liver insulin resistance and lack of compensatory beta-cell hyperplasia. *Diabetes* **49**, 1880–1889 (2000).
- Withers, D. J. *et al.* Disruption of IRS-2 causes type 2 diabetes in mice. *Nature* **391**, 900–904 (1998).
- Nakae, J. *et al.* Regulation of insulin action and pancreatic beta-cell function by mutated alleles of the gene encoding forkhead transcription factor Foxo1. *Nature Genet.* **32**, 245–253 (2002).
- Malecki, M. T. Genetics of type 2 diabetes mellitus. *Diabetes Res. Clin. Pract.* **68**, S10–S21 (2005).
- Mora, A., Lipina, C., Tronche, F., Sutherland, C. & Alessi, D. R. Deficiency of PDK1 in liver results in glucose intolerance, impairment of insulin-regulated gene expression and liver failure. *Biochem. J.* **385**, 639–648 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Franco for ARF6 and EFA6 plasmids, R. Quirion for the EGFP–FoxO1A plasmid, J. Kuriyan for the Ras and Sos plasmids, J.-L. Parent for the GST–GGA3 plasmid, M. Hoch for the steppke plasmid, V. Fieberg, K. Rotscheldt, N. Kuhn, R. Tolba, and A. Carney for technical assistance and the members of the Famulok laboratory for helpful discussions. This work was supported by grants from the Deutsche Forschungsgemeinschaft, the Sonderforschungsbereiche 645 and 704, the Fonds der Chemischen Industrie (to M.F.), and the Alexander von Humboldt foundation (to S.G.S.).

Author Contributions M. H. and A.S. contributed equally to this work. M.H. and A.S. performed and designed, with M.F., most of the included studies. I.G. performed the aptamer displacement screen and binding and *in vitro* inhibition analyses of Secins. S.G.S. synthesized all Secin derivatives. W.K. provided cytohesin and ARF expression plasmids, E.K. produced the cyh3 monoclonal antibody and B.P. characterized it. T.Q. performed the analysis of Golgi integrity and ARF6 membrane recruitment. I.B. and A.S. did the immunoprecipitation experiments. M.F. supervised the research project, and assisted in the experimental design. All authors discussed the experimental results. A.S. and M.F. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.F. (m.famulok@uni-bonn.de).

The cytohesin Steppke is essential for insulin signalling in *Drosophila*

Bernhard Fuss^{1*}, Thomas Becker^{1*}, Ingo Zinke^{1*} & Michael Hoch¹

In metazoans, the insulin signalling pathway has a key function in regulating energy metabolism and organismal growth¹. Its activation stimulates a highly conserved downstream kinase cascade that includes phosphatidylinositol-3-OH kinase (PI(3)K) and the serine–threonine protein kinase Akt. This study identifies a new component of insulin signalling in *Drosophila*, the *steppke* gene (*step*). *step* encodes a member of the cytohesin family of guanine nucleotide exchange factors (GEFs), which have been characterized as activators for ADP-ribosylation factor (ARF) GTPases^{2–4}. In *step* mutant animals both cell size and cell number are reduced, resulting in decreased body size and body weight in larvae, pupae and adults. *step* acts upstream of PI(3)K and is required for the proper regulation of Akt and the transcription factor FOXO. Temporally controlled interference with the GEF activity of the Step protein by feeding the chemical inhibitor SecinH3 causes a block of insulin signalling and a phenocopy of the *step* mutant growth defect. Step represses its own expression and the synthesis of growth inhibitors such as the translational repressor 4E-BP. Our findings indicate a crucial role of an ARF–GEF in insulin signalling that has implications for understanding insulin-related disorders, such as diabetes and obesity.

All animals coordinate growth to reach their final size and shape. The insulin–insulin-like growth factor signalling pathway, which is genetically conserved from flies to humans, has been identified as a key regulator of cell growth in response to extrinsic signals such as growth factors and nutrient availability¹. In mammals, loss of the ability to respond to insulin, a phenomenon known as insulin resistance, is associated with pathological manifestations such as type 2 diabetes^{5,6}. In *Drosophila*, activation of a unique insulin-like receptor (InR) stimulates a conserved downstream cascade that includes PI(3)K and Akt ref. 1. This signalling cascade controls organismal growth directly by regulating cell size and cell number⁷.

In a search for genes controlling larval growth in *Drosophila*, we identified a genetic locus which we named *steppke* (*step*). Molecular analysis and genetic rescue experiments show that the lethality of the P element alleles is linked to the *step* gene function (Supplementary Figs 1 and 2). The *step* gene encodes a protein that belongs to the highly conserved cytohesin protein family⁸ of GEFs, which consists of four family members in humans and one family member in invertebrates such as the nematode, mosquito and fly (Supplementary Fig. 3). GEFs mediate the exchange of GDP for GTP on the ARFs, which belong to the Ras superfamily of small GTPases⁴. Like other Ras-related GTP-binding proteins, the ARF proteins cycle between their active GTP-bound and inactive GDP-bound conformations. In concert with ARFs, cytohesin proteins regulate vesicle trafficking, cell adhesion, migration and structural organization at the cell surface^{4,9}.

Cytohesin proteins contain two characteristic motifs: a Sec7 domain responsible for the GEF activity^{2,3}, and a pleckstrin homology

domain (PH) required for plasma membrane recruitment as a result of specific binding to phosphatidylinositol-3,4,5-trisphosphate^{10,11}, the second messenger generated by class I PI(3)Ks. The Sec7 and PH domains of Step are highly conserved compared with the corresponding protein domains of mammalian cytohesins (Supplementary Fig. 3).

Phenotypic analysis of homozygous *step*^{k08110} and *step*^{SH0323} mutants and transheterozygous allelic combinations indicate an essential role of *step* in regulating growth and body size at all stages of the *Drosophila* life cycle (Fig. 1a–c). Both males and females of *step*^{k08110}/*step*^{SH0323} transheterozygous adults are significantly smaller than control animals (Fig. 1a) and they show a drastic decrease in weight (Fig. 1b); however, the body proportions of these animals are not changed. Consistently, larval and pupal development are also slowed down in *step* mutants (data not shown) and body size is reduced (Fig. 1c, and Supplementary Fig. 4). The observed growth defects mimic a starvation phenotype that is not caused by a failure of food intake, as verified by feeding coloured yeast and by the analysis of a metabolic marker gene (Fig. 2a).

It is known that larval growth is largely based on an increase in cell size in all terminally differentiated tissues, which is accomplished by endoreplication, a modified cell cycle, consisting of successive rounds of DNA synthesis without intervening mitoses¹². To examine the cause for the growth defects of *step* mutant larvae, we investigated cell cycle activity in the midgut and the salivary glands, which are representative endoreplicating tissues in the larval stage¹³. We find a general decrease in endoreplication activity, indicated by a slowing down of the S phase of the cell cycle (Fig. 1e, f). Both the size and the total number of salivary gland cells are decreased, resulting in a smaller organ (Fig. 1g, h).

Because we observe embryonic lethality in a small proportion of the homozygous *step*^{k08110} mutants, we wished to exclude the possibility that the growth defects observed in the mutant larvae derive from a defect laid down during embryogenesis. For this purpose we established an assay to analyse *step* function exclusively in the larval stage, in which the growth rate is maximal. We made use of the small molecule SecinH3, which was recently identified as an inhibitor of the *Drosophila* Step protein and the vertebrate cytohesin family members¹⁴. SecinH3 binds to the Sec7 domain of Step, thereby inhibiting the guanine nucleotide exchange of interacting ARF proteins¹⁴. Feeding SecinH3 to wild-type larvae induced a phenocopy of the growth defects observed in *step* mutants and led to a marked decrease in body size (Fig. 1d). We conclude from the phenotypic analysis of *step* mutants and from the experiments inhibiting Step protein function directly by using the chemical inhibitor SecinH3 that Step is essential for organismal growth of *Drosophila* larvae, pupae and adults.

In *step* mutants, organismal growth is strongly reduced and development is delayed, which is also a hallmark of mutants affecting the

¹LIMES-Institute, Program Unit Development and Genetics, Laboratory for Molecular Developmental Biology, University of Bonn, Meckenheimer Allee 169, D-53115 Bonn, Germany.
*These authors contributed equally to this work.

insulin signalling pathway¹. To investigate whether *step* has a function in insulin signalling, we analysed the expression of two known target genes^{15,16} of the pathway in *step* mutants, namely *4E-BP*, encoding a translational repressor¹⁷, and *InR*, encoding the insulin receptor¹⁸, by using quantitative reverse-transcriptase-mediated polymerase chain reaction (RT-PCR); both *4E-BP* and *InR* transcription are upregulated in response to repressed insulin signalling. *Lipase3* (*Lip3*) expression was used as a starvation marker in these experiments¹⁹. In *step* mutant larvae (Fig. 2a) and also in wild-type larvae treated with the Step inhibitor SecinH3 (Fig. 2b), *4E-BP* and *InR* transcription is activated, whereas *Lip3* expression is unaffected (compare with Fig. 3d). This indicates that the growth phenotype observed in *step* mutant larvae is not caused by a complete block of nutrition but is associated with a specific downregulation of insulin signalling activity. Similarly, interfering with Step function by feeding SecinH3 to transheterozygous *step* mutant flies or applying SecinH3 in S2 tissue culture cells also results in an activation of *4E-BP* and *InR* transcription (Fig. 2c, d).

It has been shown previously that *4E-BP* and *InR* are target genes of the transcription factor FOXO (for forkhead box, sub-group 'O')^{15,16}. In *Drosophila* cells, insulin receptor signalling results in a high activity of PI(3)K and phosphorylation of Akt. Akt phosphorylates FOXO and causes cytoplasmic retention of FOXO, whereas low activities of

PI(3)K and Akt allow FOXO to enter the nucleus, where it promotes the expression of factors such as *4E-BP* that retard cell growth and proliferation. In *step* mutant larvae (Fig. 3a) or in S2 tissue culture cells in which Step protein function is inhibited with SecinH3 (Supplementary Fig. 5), we find a nuclear localization of FOXO, indicating that *step* is required for insulin-signalling-dependent cytoplasmic localization of FOXO. Because this is regulated by phosphorylation by means of Akt, we tested whether *step* is necessary for Akt phosphorylation and found that under conditions in which the *step* function is affected, the amount of phosphorylated Akt protein is significantly decreased (Fig. 3b).

It has been shown that activation of Akt during growth in *Drosophila* is regulated by the class I PI(3)K Dp110 (ref. 20). Overexpression of Dp110-CAAX, a constitutively active form of PI(3)K (ref. 21), in wing or eye imaginal discs enhances cellular growth, resulting in enlarged cells and organs^{21,22}, whereas mutations in Dp110 are lethal and result in a larval growth arrest in the third instar²³. It has been shown previously that Dp110 interacts with key components of the insulin signalling pathway including Chico, PTEN and Akt to control insulin-signalling-dependent cell and organ growth in *Drosophila*^{23–25}. To test whether *step* acts together with PI(3)K in a common pathway involved in Akt and FOXO regulation and, if so, to address whether *step* is genetically upstream or downstream of PI(3)K in the insulin pathway, we expressed Dp110-CAAX (ref. 21) in heterozygous and transheterozygous *step* mutant animals.

step mutant adults are greatly decreased in size and weight in comparison with wild-type animals (Figs 1b and 3c). In control flies in which Dp110-CAAX has been overexpressed, body size and weight

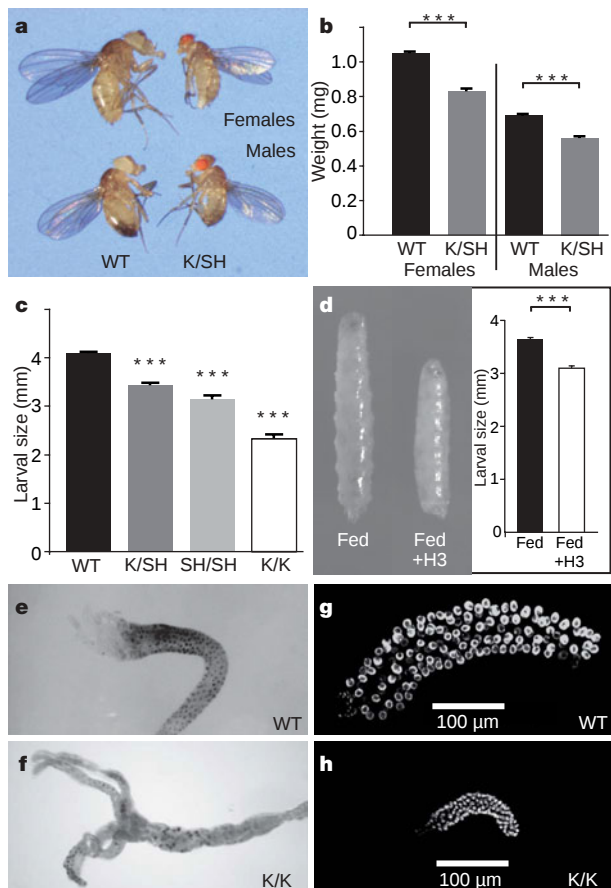


Figure 1 | *step* function is required for growth. **a**, Adult transheterozygous *step*^{K08110/step}^{SH0323} mutants (K/SH) compared with wild type (WT). **b**, Average weight of K/SH (females, $n = 28$; males, $n = 25$) and WT (females, $n = 51$; males, $n = 44$) adults. **c**, Average size of WT ($n = 141$), K/SH ($n = 140$), SH/SH ($n = 61$) and K/K ($n = 59$) larvae. **d**, Left: WT larvae fed with (Fed+H3) or without (Fed) the Step inhibitor SecinH3. Right: average size of Fed+H3 ($n = 136$) and Fed ($n = 98$) larvae. **e**, **f**, Endoreplication patterns in gut of WT (**e**) and K/K (**f**) third-instar larvae. **g**, **h**, Nuclear staining of salivary glands of WT (**g**) and K/K (**h**) third-instar larvae. Nuclei counts: WT, 151 ± 4 (mean $\pm$ s.e.m.; $n = 9$); K/K, 128 ± 4 ($n = 9$). Asterisks in **b–d** indicate statistical significance ($P < 0.001$); error bars indicate s.e.m.

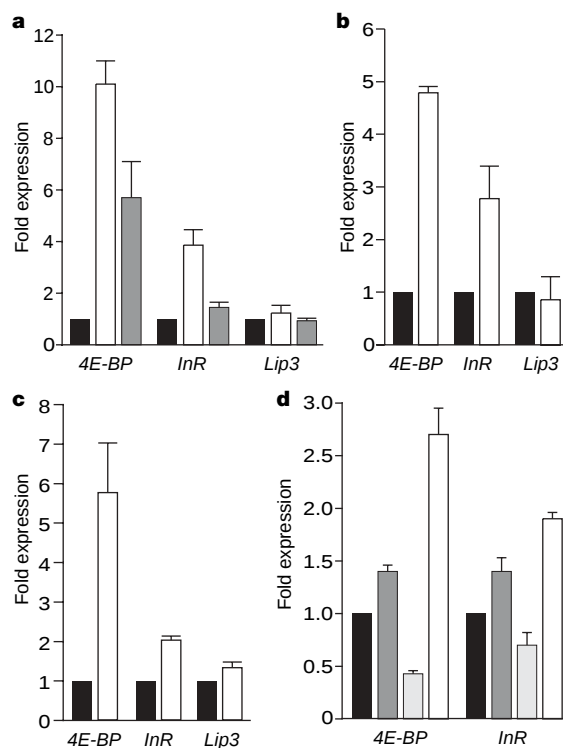


Figure 2 | Mutations in *step* and inhibition of Step function impair insulin-dependent transcription. **a**, Transcription of the genes encoding eIF-4E-binding protein (*4E-BP*) and insulin-like receptor (*InR*) and *Lip3* in wild-type (black bars), K/K (white bars) and K/SH (grey bars) first-instar larvae. **b**, *4E-BP*, *InR* and *Lip3* transcription in SecinH3-fed (white bars) or control (black bars) wild-type third-instar larvae. **c**, *4E-BP*, *InR* and *Lip3* transcription in SecinH3-fed (white bars) or control (black bars) adult K/SH flies. **d**, Reduced insulin signalling in SecinH3-treated *Drosophila* S2 cells. Black bars, no insulin and no SecinH3; dark grey bars, no insulin but SecinH3 present; light grey bars, insulin present but no SecinH3; white bars, both insulin and SecinH3 present. Error bars indicate s.e.m.

are greatly increased in comparison with wild-type flies. If *step* were positioned downstream of PI(3)K, the oversize phenotype induced by the expression of Dp110-CAAX should be suppressed or at least strongly reduced, whereas if *step* were positioned upstream of PI(3)K, Dp110-CAAX expression would rescue the growth phenotype of *step* mutants. As shown in Fig. 3c, we find the latter, providing *in vivo* evidence that the cytohesin family member *step* is upstream of PI(3)K.

Tight regulation of insulin signalling activity has been shown to be crucial for cell and organ growth in *Drosophila* and for numerous growth-related and homeostasis-related diseases such as cancer and type 2 diabetes in humans^{1,5,6}. It is known from recent studies in

Drosophila that InR represses its own synthesis by a feedback mechanism directed by the transcription factor FOXO¹⁶. To test whether *step* is also part of a negative feedback control mechanism, we analysed *step* transcription at different levels of insulin signalling activity *in vivo* by using quantitative RT-PCR experiments (Fig. 3d, e). We find that, similarly to the *4E-BP* and *InR* genes, *step* transcription is upregulated under conditions promoting FOXO activity such as starvation (Fig. 3d) or in mutants of the insulin signalling pathway, such as *chico* mutants (Supplementary Fig. 6). Consistently, *step* transcription is induced 24-fold in response to a brief pulse of ectopic FOXO expression during larval development (Fig. 3e). These results indicate a FOXO-dependent transcription of *step*, which may be direct, presumably through several FOXO consensus binding motifs^{15,26–28} present in the *step* promoter (Supplementary Fig. 7), or indirect.

We therefore propose that Step is a previously unrecognized and essential component of the insulin signalling cascade in *Drosophila* that regulates organismal growth. Our results are consistent with the findings of a parallel study¹⁴ on the role of mammalian cytohesins. Both papers provide independent evidence for the central involvement of cytohesins in the insulin pathway upstream of PI(3)K and show a functional conservation of these proteins for at least 900 million years.

METHODS

Phenotype characterization and statistics. Adult flies were measured on a Sartorius BP211D scale 7 days after eclosion. Larvae were grown on yeast and measured with an Olympus SZX12 microscope and analySIS software 96 h after egg deposition. For drug-feeding experiments, 10 mg SecinH3 was mixed with 1 g of autoclaved brewing yeast. All values are presented as means $\pm$ s.e.m. *P* values were calculated with the two-tailed Student's *t*-test for two samples with unequal variance (three asterisks, $P < 0.001$).

Immunohistochemistry and bromodeoxyuridine labelling. Early third-instar larvae were fed for 12 h with Carmin-red-stained yeast containing 5-bromodeoxyuridine at a concentration of 100 $\mu\text{g ml}^{-1}$. Antibodies used were the following: anti-bromodeoxyuridine (Becton Dickinson; 1:100 dilution), alkaline phosphatase-conjugated goat anti-mouse (Dianova; 1:500 dilution) and anti-FOXO (gift from E. Hafen; 1:1,000 dilution).

Fly strains, temperature shift and rescue experiments. Fly stocks were obtained from the Bloomington or Szeged stock centre. *UAS-foxo* flies were a gift from M. Tatar. For *hs-gal4*-dependent FOXO overexpression, early third-instar larvae were heat-shocked for 1 h at 37 °C. Rescue experiments were done with a full-length *step* complementary DNA (GenBank accession number AY071322) cloned into the pUAST-C-GFP vector.

Real time RT-PCR. RNA II kit (Macherey-Nagel) was used for RNA preparation, and QuantiTect kit (Qiagen) was used for combined genomic DNA digestion and cDNA synthesis. Quantitative PCR was performed with the iQ5 Real-Time PCR detection system (Bio-Rad). Experiments were repeated with independently isolated RNA samples from different egg collections.

Western blot analysis. Anti-(*Drosophila* phospho-Akt (Ser 505)) and anti-(*Drosophila* Akt) primary antibodies (Cell Signaling; 1:1,000 dilution; incubation was overnight in TBS with 5% BSA and 0.1% Tween-20) as well as peroxidase-coupled goat anti-rabbit (Santa Cruz; 1:15,000 dilution) secondary antibody was used, followed by enhanced chemiluminescence detection (Amersham) with a VersaDoc 5000 charge-coupled device camera and QuantityOne software (Bio-Rad).

Received 21 September; accepted 3 November 2006.

- Hafen, E. Cancer, type 2 diabetes, and ageing: news from flies and worms. *Swiss Med. Wkly* 134, 711–719 (2004).
- Jackson, C. L. & Casanova, J. E. Turning on ARF: the Sec7 family of guanine-nucleotide-exchange factors. *Trends Cell Biol.* 10, 60–67 (2000).
- Randazzo, P. A. et al. Cytohesins and centaurins: mediators of PI 3-kinase regulated Arf signalling. *Trends Biochem. Sci.* 26, 220–221 (2001).
- D'Souza-Schorey, C. & Chavrier, P. ARF proteins: roles in membrane traffic and beyond. *Nature Rev. Mol. Cell Biol.* 7, 347–358 (2006).
- Saltiel, A. R. New perspectives into the molecular pathogenesis and treatment of type 2 diabetes. *Cell* 104, 517–529 (2001).
- Virkamaki, A., Ueki, K. & Kahn, C. R. Protein-protein interaction in insulin signalling and the molecular mechanisms of insulin resistance. *J. Clin. Invest.* 103, 931–943 (1999).
- Broggiolo, W. et al. An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. *Curr. Biol.* 11, 213–221 (2001).

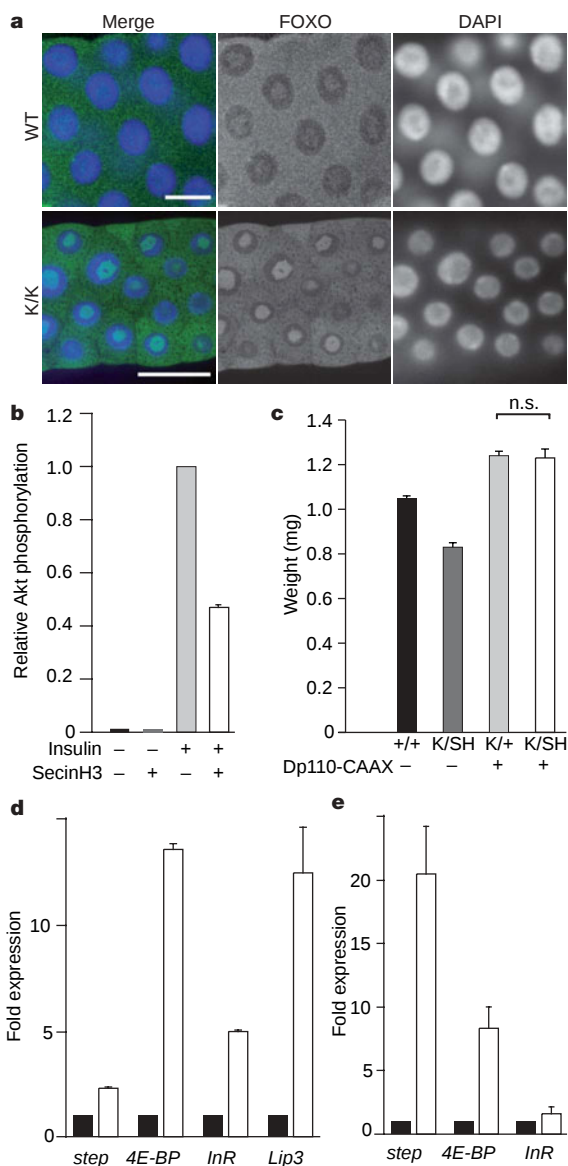


Figure 3 | Step acts upstream of PI(3)K and is involved in a negative feedback loop regulating *step* transcription. **a**, FOXO immunostaining in salivary glands of wild-type (WT) and K/K third-instar larvae. Scale bars, 25 μm . **b**, Insulin-dependent Akt phosphorylation in S2 cells treated with or without SecinH3. **c**, Average weight of female adult wild-type (+/+, $n = 79$), K/K ($n = 28$), *UAS-Dp110-CAAX*/+; *step*^{K08110}/+; *hs-gal4*/+ (Dp110-CAAX, K/+; $n = 40$) and *UAS-Dp110-CAAX*/+; *step*^{K08110}/*step*^{SH0323}; *hs-gal4*/+ (Dp110-CAAX, K/K; $n = 20$) flies. n.s., not significant ($P = 0.77$). **d**, *step*, *4E-BP*, *InR* and *Lip3* transcription in starved (white bars) or yeast-fed (black bars) second-instar wild-type larvae. **e**, *step*, *4E-BP* and *InR* transcription in wild-type (black bars) and *hs-gal4/UAS-foxo* (white bars) second-instar larvae. Error bars in **b–e** indicate s.e.m.

8. Kolanus, W. *et al.* α L β 2 integrin/LFA-1 binding to ICAM-1 induced by cytohesin-1, a cytoplasmic regulatory molecule. *Cell* **86**, 233–242 (1996).
9. Geiger, C. *et al.* Cytohesin-1 regulates β -2 integrin-mediated adhesion through both ARF-GEF function and interaction with LFA-1. *EMBO J.* **19**, 2525–2536 (2000).
10. Lietzke, S. E. *et al.* Structural basis of 3-phosphoinositide recognition by pleckstrin homology domains. *Mol. Cell* **6**, 385–394 (2000).
11. Gray, A., Van Der Kaay, J. & Downes, C. P. The pleckstrin homology domains of protein kinase B and GRP1 (general receptor for phosphoinositides-1) are sensitive and selective probes for the cellular detection of phosphatidylinositol 3,4-bisphosphate and/or phosphatidylinositol 3,4,5-trisphosphate. *Biochem. J.* **344**, 929–936 (1999).
12. Britton, J. S. & Edgar, B. A. Environmental control of the cell cycle in *Drosophila*: nutrition activates mitotic and endoreplicative cells by distinct mechanisms. *Development* **125**, 2149–2157 (1998).
13. Pierce, S. B. *et al.* *dMyc* is required for larval growth and endoreplication in *Drosophila*. *Development* **131**, 2317–2327 (2004).
14. Hafner, M. *et al.* Inhibition of cytohesins by SecinH3 leads to hepatic insulin resistance. *Nature* doi:10.1038/nature05415 (this issue).
15. Jünger, M. A. *et al.* The *Drosophila* forkhead transcription factor FOXO mediates the reduction in cell number associated with reduced insulin signalling. *J. Biol.* **2**, 20 (2003).
16. Puig, O., Marr, M. T., Ruhf, M. L. & Tjian, R. Control of cell number by *Drosophila* FOXO: downstream and feedback regulation of the insulin receptor pathway. *Genes Dev.* **17**, 2006–2020 (2003).
17. Bernal, A. & Kimbrell, D. A. *Drosophila* Thor participates in host immune defense and connects a translational regulator with innate immunity. *Proc. Natl Acad. Sci. USA* **97**, 6019–6024 (2000).
18. Chen, C., Jack, J. & Garofalo, R. S. The *Drosophila* insulin receptor is required for normal growth. *Endocrinology* **137**, 846–856 (1996).
19. Zinke, I., Schutz, C. S., Katzenberger, J. D., Bauer, M. & Pankratz, M. J. Nutrient control of gene expression in *Drosophila*: microarray analysis of starvation and sugar-dependent response. *EMBO J.* **21**, 6162–6173 (2002).
20. Verdu, J., Buratovich, M. A., Wilder, E. L. & Birnbaum, M. J. Cell-autonomous regulation of cell and organ growth in *Drosophila* by Akt/PKB. *Nature Cell Biol.* **1**, 500–506 (1999).
21. Leivers, S. J., Weinkove, D., MacDougall, L. K., Hafen, E. & Waterfield, M. D. The *Drosophila* phosphoinositide 3-kinase Dp110 promotes cell growth. *EMBO J.* **15**, 6584–6594 (1996).
22. Goberdhan, D. C., Paricio, N., Goodman, E. C., Mlodzik, M. & Wilson, C. *Drosophila* tumor suppressor PTEN controls cell size and number by antagonizing the Chico/PI3-kinase signaling pathway. *Genes Dev.* **13**, 3244–3258 (1999).
23. Weinkove, D., Neufeld, T. P., Twardzik, T., Waterfield, M. D. & Leivers, S. J. Regulation of imaginal disc cell size, cell number and organ size by *Drosophila* class I(A) phosphoinositide 3-kinase and its adaptor. *Curr. Biol.* **9**, 1019–1029 (1999).
24. Gao, X., Neufeld, T. P. & Pan, D. *Drosophila* PTEN regulates cell growth and proliferation through PI3K-dependent and -independent pathways. *Dev. Biol.* **221**, 404–418 (2000).
25. Stocker, H. *et al.* Living with lethal PIP3 levels: viability of flies lacking PTEN restored by a PH domain mutation in Akt/PKB. *Science* **295**, 2088–2091 (2002).
26. Furuyama, T., Nakazawa, T., Nakano, I. & Mori, N. Identification of the differential distribution patterns of mRNAs and consensus binding sequences for mouse DAF-16 homologues. *Biochem. J.* **349**, 629–634 (2000).
27. Xuan, Z. & Zhang, M. Q. From worm to human: bioinformatics approaches to identify FOXO target genes. *Mech. Ageing Dev.* **126**, 209–215 (2005).
28. Kaufmann, E. & Knochel, W. Five years on the wings of fork head. *Mech. Dev.* **57**, 3–20 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank W. Kolanus for valuable discussions on cytohesins; M. Famulok for providing the Step inhibitor SecinH3 and for sharing unpublished results; S. Cohen and M. Günther Behr for reagents; A. Schmitz for help with pAkt quantifications; C. Müller for technical assistance; and the members of the Famulok and Hoch laboratories for helpful comments. This work was supported by grants from the Deutsche Forschungsgemeinschaft, the Sonderforschungsbereiche 645 (to M.H.) and 704 (to M.H. and B.F.).

Author Contributions T.B. and I.Z. performed the growth and body size analysis in *step* mutants. T.B. analysed the insulin signalling activity in *step* and *chico* mutants and under starvation using quantitative RT-PCR. B.F. performed the cell cycle analysis, cell size analysis in *step* mutants, pAkt quantifications, *in vitro* and *in vivo* FOXO localization studies, *in vitro* growth inhibition analysis, *step* enhancer analysis and ectopic FOXO expression experiments. I.Z. generated the *in vivo* chemical genetics data and the Dp110-CAAX rescue experiment. I.Z. and B.F. contributed to the *step* rescue experiment. M.H. supervised the research project, and assisted in the experimental design. All authors discussed the experimental results. B.F. and M.H. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.H. (m.hoch@uni-bonn.de).

Long-lasting arrest of murine polycystic kidney disease with CDK inhibitor roscovitine

Nikolay O. Bukanov¹, Laurie A. Smith¹, Katherine W. Klinger², Steven R. Ledbetter¹
& Oxana Ibraghimov-Beskrovnya¹

Polycystic kidney diseases (PKDs) are primarily characterized by the growth of fluid-filled cysts in renal tubules leading to end-stage renal disease^{1–3}. Mutations in the *PKD1* or *PKD2* genes lead to autosomal dominant PKD (ADPKD), a slowly developing adult form^{4,5}. Autosomal recessive polycystic kidney disease results from mutations in the *PKHD1* gene, affects newborn infants and progresses very rapidly^{6,7}. No effective treatment is currently available for PKD. A previously unrecognized site of subcellular localization was recently discovered for all proteins known to be disrupted in PKD: primary cilia^{8,9}. Because ciliary functions seem to be involved in cell cycle regulation, disruption of proteins associated with cilia or centrioles may directly affect the cell cycle and proliferation, resulting in cystic disease^{10–12}. We therefore reasoned that the dysregulated cell cycle may be the most proximal cause of cystogenesis, and that intervention targeted at this point could provide significant therapeutic benefit for PKD. Here we show that treatment with the cyclin-dependent kinase (CDK) inhibitor (*R*)-roscovitine does indeed yield effective arrest of cystic disease in *jck* and *cpk* mouse models of PKD. Continuous daily administration of the drug is not required to achieve efficacy; pulse treatment provides a robust, long-lasting effect, indicating potential clinical benefits for a lifelong therapy. Molecular studies of the mechanism of action reveal effective cell-cycle arrest, transcriptional inhibition and attenuation of apoptosis. We found that roscovitine is active against cysts originating from different parts of the nephron, a desirable feature for the treatment of ADPKD, in which cysts form in multiple nephron segments. Our results indicate that inhibition of CDK is a new and effective approach to the treatment of PKD.

To determine the potential of CDK inhibition as a new therapeutic approach for the treatment of PKD we employed (*R*)-roscovitine (CYC202), a potent inhibitor of Cdk2–cyclin E with a 50% inhibitory concentration (IC₅₀) of 0.1 μM as well as Cdk7–cyclin H (IC₅₀ 0.4 μM), Cdk9–cyclin T1 (IC₅₀ 0.8 μM) and Cdk5–p35–p25 (IC₅₀ 0.16 μM)^{13,14}. It is currently in clinical trials as an anticancer agent^{15,16}. We tested the effect of roscovitine on slowly progressive renal cystic disease in *jck* mice. The development of PKD in the *jck* mice resembles human disease in many ways, as described previously¹⁷. Thus, similarly to ADPKD, *jck* mice develop cysts in multiple nephron segments and display gender dimorphism with more aggressive disease in males. *jck/jck* males were injected daily with 50 and 150 mg kg^{−1} roscovitine from day 26 to day 64 for a total of 5 weeks (Fig. 1a, b). The drug was well tolerated with no weight loss, mortality or toxicity (Supplementary Table 1). This is not surprising because roscovitine seems to be well tolerated in preclinical mouse models up to 2,000 mg kg^{−1} orally and in phase I/II clinical trials in solid tumours¹⁶. The roscovitine-treated group showed a significant decrease in the ratio of kidney weight to body weight and in cystic

volumes, in a dose-dependent manner (Fig. 1a, b, and Supplementary Table 1). In addition, blood urea nitrogen (BUN) was significantly decreased in the 50 mg kg^{−1} roscovitine-treated group and normalized with 150 mg kg^{−1} roscovitine (Fig. 1a). Roscovitine also effectively inhibited PKD in *jck/jck* females, in which the disease is less severe than in males (Supplementary Table 1).

Given the prominent inhibition of cystogenesis in *jck* mice with daily dosing of roscovitine for 5 weeks, we examined whether continuous administration of the drug is required for efficacy, or whether pulse therapy with roscovitine will produce a long-lasting effect. We administered 150 mg kg^{−1} roscovitine daily to *jck/jck* males for 3 weeks, followed by 2 weeks without treatment (Fig. 1c; 3/2 schedule). A long-lasting anti-cystic effect was detected with normalized BUN levels. We also employed a different schedule in which the drug was administered daily for 1 week on treatment followed by 1 week off treatment for a total of 5 weeks (Fig. 1c; 1/1 schedule). Again, a significant decrease in cystogenesis was detected, whereas the decrease in BUN was not statistically significant. Roscovitine therefore shows a long-lasting treatment effect after drug withdrawal and is also efficacious with intermittent dosing. This potential for pulse therapy may prove extremely important for the treatment of human PKD and other chronic kidney diseases for which lifelong safe therapy is required.

Because different animal models each manifest different facets of human PKD, the efficacy of a therapy needs to be assessed in more than one model. We have analysed the effect of roscovitine on aggressive PKD in *cpk* mice. Daily treatment of *cpk* mice with 100 mg kg^{−1} roscovitine from day 7 to day 21 resulted in a significant decrease in PKD (Fig. 2 and Supplementary Table 2). Thus, roscovitine effectively inhibited renal cystic disease in two models of PKD, namely slowly progressive *jck* and aggressive *cpk*.

It is possible that cysts originating from functionally different nephron segments may respond differently to therapy. We have previously shown that similarly to ADPKD, *jck* mice develop cysts in multiple parts of the nephron¹⁷. We show that roscovitine is highly effective against cysts of different origins: the cortical collecting duct (CD), the distal tubule (DT) and the loop of Henle (LH) (Supplementary Fig. S2).

To understand the mode of action of roscovitine in PKD, we analysed its effect on the proliferation of cystic epithelia. First, with proliferating cell nuclear antigen (PCNA) immunostaining we confirmed high proliferation rates for human ADPKD cyst-lining epithelia (Fig. 3a). Similarly, a large number of PCNA-positive cells were seen in *jck* cysts (Fig. 3b). In contrast, a quantitative decrease in PCNA-positive cells was detected in roscovitine-treated *jck* kidney (Fig. 3b, c). Next we examined the effects of roscovitine treatment on the cell cycle. Two kinases, Cdk4/6–cyclin D and Cdk2–cyclin E, and the transcription complex that includes retinoblastoma protein (Rb)

¹Cell Biology, ²Genetics and Genomics, Genzyme Corporation, 5 Mountain Road Framingham, Massachusetts 01701-9322, USA.

are pivotal in controlling the G1/S checkpoint (Fig. 3e). Because a primary target of roscovitine is the Cdk2–cyclin E complex that phosphorylates and inactivates Rb tumour suppressor, we examined Rb phosphorylation (Rb-p) in wild-type (WT), *jck* and roscovitine-treated kidneys (Fig. 3d, e). The Rb-p form was significantly increased in cystic kidneys. Treatment with roscovitine resulted in the dephosphorylation of Rb, suggesting an effective G1/S cell-cycle block. In addition, treatment with roscovitine normalized the levels of the Rb-associated proteins Rb2 and retinoblastoma-binding protein (Fig. 3d). Although the level of cyclin D1 was not substantially increased in cystic kidneys, its phosphorylation (at Thr 286) was significantly decreased (Fig. 3d). Inhibition of cyclin D1 phosphor-

ylation in the *jck* kidney probably prevents its degradation through the ubiquitin–proteasome pathway¹⁸. Roscovitine normalized the phosphorylation level of cyclin D1 (Fig. 3d). The levels of cyclin D2 and D3 expression were increased in cystic kidneys and down-regulated by roscovitine. Roscovitine also affected signalling from Ras–Raf to MEK–ERKs, known to regulate cyclin D1 by decreasing the level of ERK 1/2 activation (Fig. 3d). This effect is likely to be mediated by direct targeting of ERK 2 by roscovitine¹⁹. The importance of MEK–ERK signalling in PKD has been confirmed by showing that MEK inhibition slows cystogenesis *in vivo*²⁰.

Roscovitine also inhibits Cdk7 and Cdk9, which regulate transcription by phosphorylating the carboxy-terminal domain of RNA polymerase II (RNA pol II). Inhibition of RNA pol II-dependent transcription may generate additional benefits in slowing cystogenesis, as described previously for cancer¹⁵. Moderately increased levels of Cdk7 and Cdk9 detected in cystic kidneys resulted in greatly increased phosphorylation of RNA pol II (Fig. 3d). In contrast, roscovitine-treated samples were almost indistinguishable from wild-type controls, showing markedly decreased levels of RNA pol II phosphorylation. Thus, roscovitine attenuates cystogenesis through cell-cycle arrest and transcriptional inhibition. Roscovitine also targets Cdk5, a unique kinase activated by complexing with activator molecules p35 and mainly its cleaved product p25 (ref. 13). Abnormal Cdk5 activity has been shown to have a function in neurodegenerative diseases, and cleavage of p35 to p25 by calpain has been linked to activation of Cdk5 and cell death²¹. We detected increased Cdk5 expression in *jck* kidneys in comparison with wild-type controls, accompanied by the generation of p25 cleaved product, whereas p35 was predominantly seen in wild-type kidney (Fig. 3d). Treatment with roscovitine resulted in the downregulation of Cdk5–p25 and normalized levels of p35 (Fig. 3d). It is possible that targeting Cdk5 with roscovitine may affect apoptotic pathways in cystic disease.

Apoptosis is causally linked to cystogenesis: deletion of anti-apoptotic genes encoding Bcl-2 and AP-2 β and overexpression of pro-apoptotic c-Myc in mice results in renal cystic disease^{22,23}. Treatment with roscovitine decreased the number of TdT-mediated dUTP nick

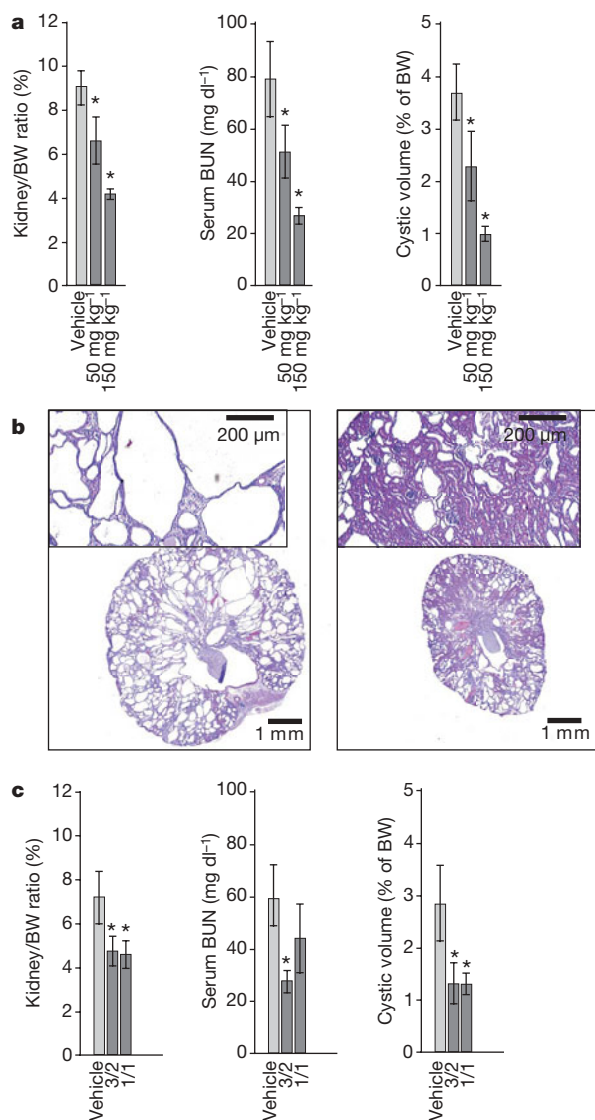


Figure 1 | Roscovitine inhibits PKD in *jck* mice. **a**, Dose-dependent effect of roscovitine on disease progression in *jck* males (50 or 150 mg kg⁻¹ daily, starting from day 26 for a total of 5 weeks). Asterisk, $P < 0.05$ compared with vehicle controls. Error bars indicate s.d. (17, 8 and 9 mice were used for the groups treated with vehicle and with 50 mg kg⁻¹ and 150 mg kg⁻¹ roscovitine, respectively). BW, body weight; BUN, blood urea nitrogen. **b**, Representative kidney sections from *jck* mice treated with vehicle control (left) and with roscovitine (right). **c**, Long-lasting effect of roscovitine treatment in *jck* mice. *Jck* mice were treated with 150 mg kg⁻¹ of roscovitine starting at day 26 either for 3 weeks daily followed by 2 weeks off treatment (3/2 schedule) or by alternating 1 week on treatment with 1 week off treatment (1/1 schedule). Asterisk, $P < 0.05$ compared with vehicle controls. Error bars indicate $\pm$ s.d. (6, 5 and 9 mice were used for vehicle, 3/2 and 1/1 groups, respectively).

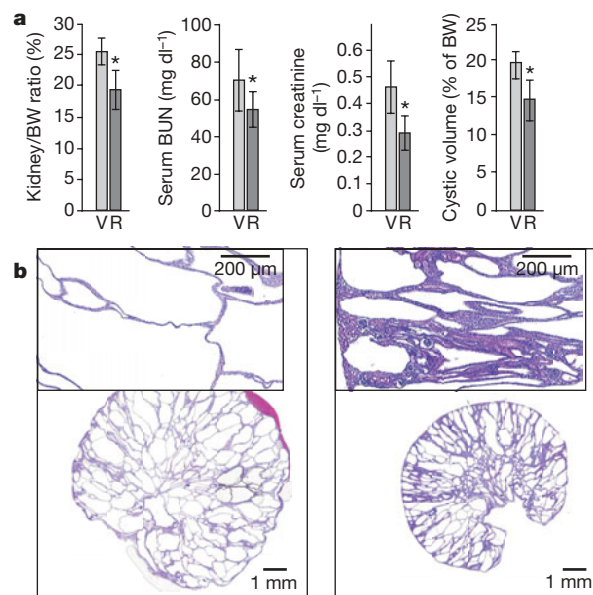


Figure 2 | Roscovitine effectively attenuates disease progression in *cpk* mice. **a**, Roscovitine (R, 100 mg kg⁻¹) or vehicle (V) was administered daily from postnatal day 7 for a total of 2 weeks. Asterisk, $P < 0.05$ compared with control. Error bars indicate s.d. (23 and 20 mice were used for the groups treated with vehicle and with roscovitine, respectively). **b**, Representative kidney sections from *cpk* mice treated with vehicle (left) and roscovitine (right).

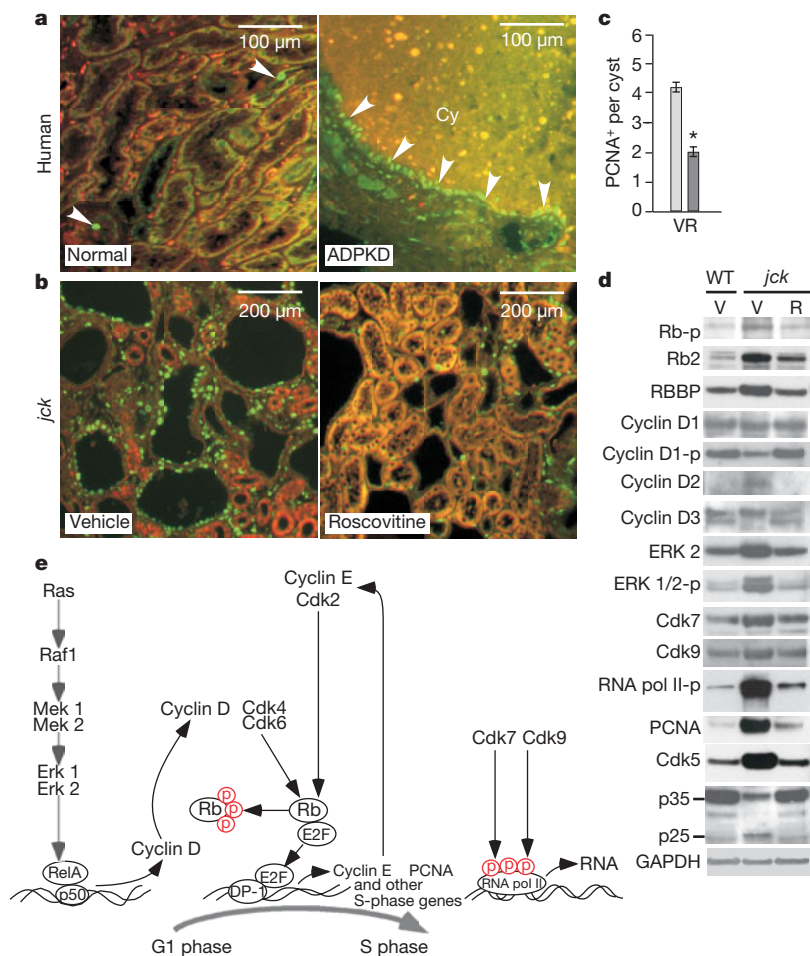


Figure 3 | Cell cycle blockade in *jck* kidney with roscovitine. **a**, Increased proliferation of human ADPKD epithelia (PCNA, green) is indicated by arrows. Cy, cystic lumen. **b**, Roscovitine inhibits the proliferation of cystic epithelial cells in *jck* kidney (indicated by PCNA staining (green)). **c**, MetaMorph quantification of the anti-proliferative effect of roscovitine in vehicle (V)-treated and roscovitine (R)-treated animals expressed as average number of PCNA-positive (PCNA(+)) cells per cyst. Error bars indicate s.e.m. (450–500 individual cysts were analysed for each group). **d**, Cell-cycle machinery in kidneys of wild-type (WT), *jck* vehicle-treated (V) (centre) and *jck* roscovitine-treated (R) mice (150 mg kg⁻¹, 5 weeks). GAPDH, glyceraldehyde-3-phosphate dehydrogenase; RBBP, retinoblastoma-binding protein. **e**, The G1/S transition in the cell cycle is activated by Rb phosphorylation and the release of E2F transcription factor by Cdk4/6–cyclin D1 and Cdk2–cyclin E. Signalling from Ras/Raf to MEK/ERKs induces cyclin D1 expression, allowing cyclin D1 to complex with Cdk4 and Cdk6. Cdk7 and Cdk9 activate RNA pol II.

end labelling (TUNEL)-positive cells, suggesting a potent anti-apoptotic effect (Fig. 4a). Caspase-2, a key initiator of the mitochondrial pathway of apoptosis, was found to be greatly induced in *jck* kidney (Fig. 4b). The increased level of ApaF1 activated by mitochondrial protein release in response to apoptotic signals was detected in *jck* kidneys (Fig. 4b). The expression of caspase-2 and ApaF1 in roscovitine-treated *jck* kidneys was indistinguishable from that of wild-type controls. Furthermore, because ApaF1 activates the caspase cascade downstream of caspase-9, we found inactive caspase-3 to be down-regulated in *jck* kidneys but restored by treatment with roscovitine (Fig. 4b). Bcl-2 and Bcl-xL proteins were upregulated with treatment, suggesting effective inhibition of apoptosis. It is possible that Cdk5 inhibition by roscovitine might have a function in its blockade of apoptosis in cystic kidneys. Such a mechanism might be responsible

for anti-apoptotic effects of roscovitine in neurodegenerative diseases^{24,25}. This activity is different from roscovitine action in tumour cells, in which it causes induction of apoptosis^{15,26}. Roscovitine treatment therefore resulted in effective reduction of apoptosis in cystic kidneys. Recently, (R)-roscovitine was shown to interact with pyridoxal kinase¹⁹. We show that pyridoxal kinase is not responsible for anti-cystic activity of roscovitine (Supplementary Fig. S5). Although the molecular basis for the long-lasting therapeutic effect of roscovitine remains to be explained, it is possible that restoration of the cell-cycle–apoptosis balance promotes a more normal epithelial phenotype. The observed effects of roscovitine on cystogenesis are probably mediated by a combination of its targets, including the possible regulation of calcium signalling, making the analysis of each molecular target an aim of future studies. Although no effective PKD therapy is

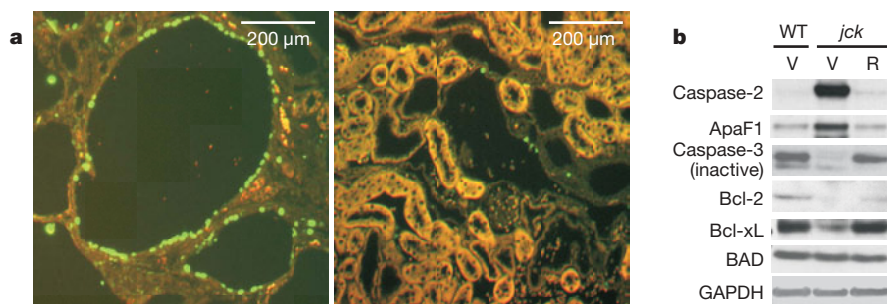


Figure 4 | Treatment with roscovitine results in decreased apoptosis in *jck* kidney. **a**, TUNEL staining (green) shows apoptotic cells in *jck* kidney (vehicle, left) mainly localized to cystic epithelia. The number of TUNEL-positive cells is greatly reduced in roscovitine-treated *jck* kidney (right).

b, Representative western blot analysis of proteins regulating apoptosis in the kidneys of wild-type (WT) vehicle (V)-treated, *jck* vehicle-treated and *jck* roscovitine (R)-treated mice. GAPDH staining shows equal protein loading (30 µg).

currently available, new therapeutic approaches are beginning to emerge^{27–30}. Multiple treatment options are needed to target PKD, and it is likely that combination therapies may prove most effective.

Our study shows for the first time the therapeutic potential for cell-cycle inhibition for the treatment of polycystic kidney diseases. Roscovitine is a selective inhibitor for CDKs with minimal off-target kinase activities¹⁹, making it a clinical candidate for ADPKD. We show that roscovitine effectively inhibited cystogenesis and improved renal function in *jck* and *cpk* models of PKD, with a long-lasting effect. Analysis of the molecular targets of roscovitine in cystogenesis showed that it acts through blockade of the cell cycle, transcriptional regulation and inhibition of apoptosis.

METHODS

Jck and cpk mouse models of PKD and experimental design. C57BL/6J *jck* + and C57BL/6J *cpk* + mice were obtained from the Jackson Laboratory and were used to establish a breeding colony maintained at Biomedical Research Models and Genzyme. (R)-Roscovitine (6-benzylamino-2-((R)-1-ethyl-2-hydroxyethylamino)-9-isopropylpurine) was obtained from A.G. Scientific. Cystic *jck/jck* mice were identified by a custom genotyping assay as described previously¹⁷; they were divided into control and treated groups and injected intraperitoneally with either vehicle or drug at specified concentrations. Details are given in Supplementary Methods.

Morphometric analysis of cystogenesis. Longitudinal and transverse kidney sections (4 µm) were stained with haematoxylin and eosin, slides were digitized with an ACIS system (Clariant) and processed with the MetaMorph Imaging Series software (Molecular Devices Corp.) and Scion Image software (Scion Corp.). Cystic percentage was measured as a ratio of the cystic area to the total section area and was used in calculations of cystic volumes (percentage of body weight) as described previously³⁰.

Immunohistochemical, immunofluorescence and western blot analyses. Detailed protocols are given in Supplementary Methods.

Statistical analysis. Data are expressed as means ± s.d. Comparisons were made by two-tailed *t*-test and significance was accepted at the 0.05 level of probability (*P* < 0.05).

Received 24 July; accepted 13 October 2006.

Published online 22 November 2006.

1. Welling, L. W. & Grantham, J. J. in *Renal Pathology* (eds Tisher, C. & Brenner, B.) 1828–1863 (Lippincott, Philadelphia, 1996).
2. Ecker, T. & Schrier, R. W. Hypertension and left ventricular hypertrophy in autosomal dominant polycystic kidney disease. *Expert Rev. Cardiovasc. Ther.* **2**, 369–374 (2004).
3. Wilson, P. D. Polycystic kidney disease: new understanding in the pathogenesis. *Int. J. Biochem. Cell Biol.* **36**, 1868–1873 (2004).
4. Hughes, J. et al. The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. *Nature Genet.* **10**, 151–160 (1995).
5. Mochizuki, T. et al. PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science* **272**, 1339–1342 (1996).
6. Onuchic, L. F. et al. PKHD1, the polycystic kidney and hepatic disease 1 gene, encodes a novel large protein containing multiple immunoglobulin-like plexin-transcription-factor domains and parallel β-helix 1 repeats. *Am. J. Hum. Genet.* **70**, 1305–1317 (2002).
7. Ward, C. J. et al. The gene mutated in autosomal recessive polycystic kidney disease encodes a large, receptor-like protein. *Nature Genet.* **30**, 259–269 (2002).
8. Pazour, G. J. Intraflagellar transport and cilia-dependent renal disease: the ciliary hypothesis of polycystic kidney disease. *J. Am. Soc. Nephrol.* **15**, 2528–2536 (2004).
9. Pan, J., Wang, Q. & Snell, W. J. Cilium-generated signaling and cilia-related disorders. *Lab. Invest.* **85**, 452–463 (2005).

10. Bhunia, A. K. et al. PKD1 induces p21^{waf1} and regulation of the cell cycle via direct activation of the JAK-STAT signaling pathway in a process requiring PKD2. *Cell* **109**, 157–168 (2002).
11. Quarumby, L. M. & Parker, J. D. Cilia and the cell cycle? *J. Cell Biol.* **169**, 707–710 (2005).
12. Woo, D. D., Miao, S. Y., Pelayo, J. C. & Woolf, A. S. Taxol inhibits progression of congenital polycystic kidney disease. *Nature* **368**, 750–753 (1994).
13. Meijer, L. et al. Biochemical and cellular effects of roscovitine, a potent and selective inhibitor of the cyclin-dependent kinases cdc2, cdk2 and cdk5. *Eur. J. Biochem.* **243**, 527–536 (1997).
14. McClue, S. J. et al. In vitro and in vivo antitumor properties of the cyclin dependent kinase inhibitor CYC202 (R-roscovitine). *Int. J. Cancer* **102**, 463–468 (2002).
15. MacCallum, D. E. et al. Seliciclib (CYC202, R-Roscovitine) induces cell death in multiple myeloma cells by inhibition of RNA polymerase II-dependent transcription and down-regulation of Mcl-1. *Cancer Res.* **65**, 5399–5407 (2005).
16. Guzi, T. CYC-202 Cyclacel. *Curr. Opin. Investig. Drugs* **5**, 1311–1318 (2004).
17. Smith, L. A. et al. Development of polycystic kidney disease in juvenile cystic kidney mice: insights into pathogenesis, ciliary abnormalities, and common features with human disease. *J. Am. Soc. Nephrol.* **17**, 2821–2831 (2006).
18. Diehl, J. A., Zindy, F. & Sherr, C. J. Inhibition of cyclin D1 phosphorylation on threonine-286 prevents its rapid degradation via the ubiquitin-proteasome pathway. *Genes Dev.* **11**, 957–972 (1997).
19. Bach, S. et al. Roscovitine targets, protein kinases and pyridoxal kinase. *J. Biol. Chem.* **280**, 31208–31219 (2005).
20. Omori, S. et al. Extracellular signal-regulated kinase inhibition slows disease progression in mice with polycystic kidney disease. *J. Am. Soc. Nephrol.* **17**, 1604–1614 (2006).
21. Lin, L., Ye, Y. & Zakeri, Z. p53, Apaf-1, caspase-3, and -9 are dispensable for Cdk5 activation during cell death. *Cell Death Differ.* **13**, 141–150 (2006).
22. Ecker, T. et al. Caspases, Bcl-2 proteins and apoptosis in autosomal-dominant polycystic kidney disease. *Kidney Int.* **61**, 1220–1230 (2002).
23. Tao, Y. et al. Caspase inhibition reduces tubular apoptosis and proliferation and slows disease progression in polycystic kidney disease. *Proc. Natl Acad. Sci. USA* **102**, 6954–6959 (2005).
24. Di Giovanni, S. et al. Cell cycle inhibition provides neuroprotection and reduces glial proliferation and scar formation after traumatic brain injury. *Proc. Natl Acad. Sci. USA* **102**, 8333–8338 (2005).
25. Verdager, E. et al. Antiapoptotic effects of roscovitine in cerebellar granule cells deprived of serum and potassium: a cell cycle-related mechanism. *Neurochem. Int.* **44**, 251–261 (2004).
26. Tirado, O. M., Mateo-Lozano, S. & Notario, V. Roscovitine is an effective inducer of apoptosis of Ewing's sarcoma family tumor cells *in vitro* and *in vivo*. *Cancer Res.* **65**, 9320–9327 (2005).
27. Sweeney, W. E. et al. Combination treatment of PKD utilizing dual inhibition of EGF-receptor activity and ligand bioavailability. *Kidney Int.* **64**, 1310–1319 (2003).
28. Tao, Y., Kim, J., Schrier, R. W. & Edelstein, C. L. Rapamycin markedly slows disease progression in a rat model of polycystic kidney disease. *J. Am. Soc. Nephrol.* **16**, 46–51 (2005).
29. Shillingford, J. M. et al. The mTOR pathway is regulated by polycystin-1, and its inhibition reverses renal cystogenesis in polycystic kidney disease. *Proc. Natl Acad. Sci. USA* **103**, 5466–5471 (2006).
30. Torres, V. E. et al. Effective treatment of an orthologous model of autosomal dominant polycystic kidney disease. *Nature Med.* **10**, 363–364 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Russo, T. Barry and A. Taylor for expert technical assistance, the staff of the Genzyme Department of Comparative Medicine and Histology unit for help with *in vivo* studies and sample preparations. We are grateful to L. Meijer and H. Galons (CNRS, Roscoff) for providing N6-methyl-(R)-roscovitine. We thank A. Smith, R. Gregory, T. Natoli, H. Husson and J. Leonard for helpful discussions and comments on this manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to O.I.-B. (oxana.beskovnaya@genzyme.com).

Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease

Corey R. Mandel¹, Syuzo Kaneko^{1*}, Hailong Zhang^{1*†}, Damara Gebauer^{1†}, Vasupradha Vethantham¹, James L. Manley¹ & Liang Tong¹

Most eukaryotic messenger RNA precursors (pre-mRNAs) undergo extensive maturational processing, including cleavage and polyadenylation at the 3'-end^{1–8}. Despite the characterization of many proteins that are required for the cleavage reaction, the identity of the endonuclease is not known^{4,9,10}. Recent analyses indicated that the 73-kDa subunit of cleavage and polyadenylation specificity factor (CPSF-73) might be the endonuclease for this and related reactions^{10–15}, although no direct data confirmed this. Here we report the crystal structures of human CPSF-73 at 2.1 Å resolution, complexed with zinc ions and a sulphate that might mimic the phosphate group of the substrate, and the related yeast protein CPSF-100 (Ydh1) at 2.5 Å resolution. Both CPSF-73 and CPSF-100 contain two domains, a metallo-β-lactamase domain and a novel β-CASP (named for metallo-β-lactamase, CPSF, Artemis, Snm1, Pso2) domain¹². The active site of CPSF-73, with two zinc ions, is located at the interface of the two domains. Purified recombinant CPSF-73 possesses RNA endonuclease activity, and mutations that disrupt zinc binding in the active site abolish this activity. Our studies provide the first direct experimental evidence that CPSF-73 is the pre-mRNA 3'-end-processing endonuclease.

CPSF-73 belongs to the metallo-β-lactamase superfamily of zinc-dependent hydrolases^{11,12}. Canonical metallo-β-lactamases contain five signature sequence motifs—Asp (motif 1), His-X-His-X-Asp-His (motif 2), His (motif 3), Asp (motif 4) and His (motif 5), most of which are ligands to the two zinc ions in their active site. Sequence conservation between CPSF-73 and the canonical metallo-β-lactamases is limited to these signature motifs. Whereas the first four motifs have been identified in the N-terminal segment of CPSF-73 (see Supplementary Fig. 1a, Supplementary Table 1), the fifth motif was uncertain, with three candidates, A (Asp or Glu), B (His) and C (His) (see Supplementary Fig. 1a), in the so-called β-CASP motif¹². Motif B was proposed to be equivalent to motif 5 in the canonical metallo-β-lactamases. Another subunit of CPSF, CPSF-100, shares sequence conservation (see Supplementary Fig. 1b) and a similar domain architecture (see Supplementary Fig. 1a) with CPSF-73 but lacks the putative zinc-binding residues.

To understand the roles of CPSF-73 and CPSF-100 in pre-mRNA 3'-end processing, we determined the structures of human CPSF-73 (residues 1–460) and yeast CPSF-100 (residues 1–720) (the crystallographic data are summarized in Supplementary Table 2). The two structures obtained for CPSF-73 were crystallized in the absence or presence of 0.5 mM zinc (although both structures contained zinc atoms; see below). We discovered serendipitously that *in situ* proteolysis by a fungal protease is crucial for the crystallization of yeast CPSF-100 (ref. 16).

The structure of CPSF-73 can be divided into two domains (Fig. 1a). The amino-terminal residues (amino acids 1–208) form a

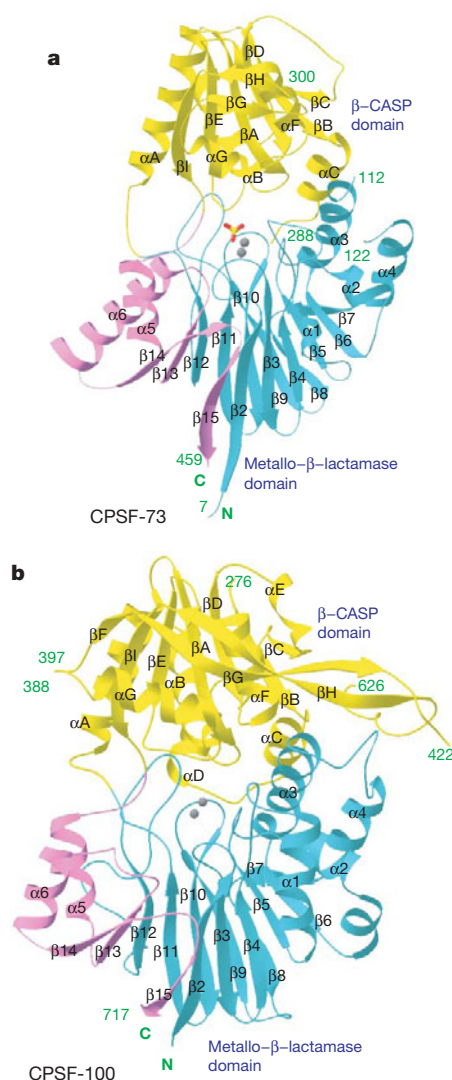


Figure 1 | Structures of human CPSF-73 and yeast CPSF-100 (Ydh1). **a**, Schematic representation of the structure of human CPSF-73. The β-strands and α-helices are labelled, and the two zinc atoms in the active site are shown as grey spheres. The sulphate ion is shown as a stick model. **b**, Schematic representation of the structure of yeast CPSF-100. The zinc atoms in the CPSF-73 structure are shown for reference. See Supplementary Fig. 2 For Schematic drawings of the metallo-β-lactamase domains of the two proteins.

¹Department of Biological Sciences, Columbia University, New York, New York 10027, USA. [†]Present addresses: Array BioPharma Inc., Boulder, Colorado 80301, USA (H.Z.); Johnson & Johnson Pharmaceutical Research & Development LLC, San Diego, California 92121, USA (D.G.).

*These authors contributed equally to this work.

domain similar to the structure of canonical metallo- β -lactamases, with a four-layered $\alpha\beta/\beta\alpha$ architecture (see Supplementary Fig. 2a). A close structural homologue of this domain is the L1 metallo- β -lactamase from *Stenotrophomonas maltophilia* (see Supplementary Fig. 2c)¹⁷, with a root mean squared (r.m.s.) distance of 3.3 Å for 167 equivalent C α atoms, calculated using the program Dali¹⁸. However, the sequence identity for these structurally equivalent residues is only 17%.

Despite the overall similarity to canonical metallo- β -lactamases, the structure of this domain of CPSF-73 contains several distinctive features. Most remarkably, residues 395–460, at the carboxy-terminal end of the expression construct (see Supplementary Fig. 1a), are also part of the metallo- β -lactamase domain (Fig. 1a). These residues add three strands (β 13 to β 15) to the central β -sandwich, and the last residue of the structure is located next to the first residue (see Supplementary Fig. 2a). These additional β -strands in the metallo- β -lactamase fold are also seen in the structure of CPSF-100 (Fig. 1b, Supplementary Fig. 2b; see below) and RNase Z (see Supplementary Fig. 2d)^{19,20}, indicating that they might be unique to RNA/DNA-processing nucleases. This feature of the metallo- β -lactamase fold is crucial for the activity of these proteins, as the loop after strand β 13 provides one of the ligands (motif C) to the zinc ions in the active site (see Supplementary Fig. 2a and see below). The r.m.s. distance for 210 equivalent C α atoms between CPSF-73 and RNase Z is 2.6 Å, but the sequence identity is only 18%. RNase Z is also distinct from CPSF-73 in that it lacks the second domain (Fig. 1a, Supplementary Fig. 2d).

The second domain of CPSF-73 covers residues 209–394 (Supplementary Fig. 1a), and contains a central, fully parallel, six-stranded β -sheet that is surrounded by five α -helices on both faces (Fig. 2a). In addition, a small β -hairpin structure is on the surface of the domain. This second domain can be considered as a cassette inserted into the metallo- β -lactamase domain, between strand β 12 and helix α 5 (Fig. 1a). As it is related to the β -CASP motif identified earlier on the basis of sequence analyses¹², we have named it the β -CASP domain (see Supplementary Fig. 1a).

The closest structural homologue of the β -CASP domain is the nucleotide-binding fold (NBF, Supplementary Fig. 3), with a Z score of 6 from Dali¹⁸. However, the β -CASP domain lacks the Walker A motif for binding nucleotide phosphate groups²¹ (see Supplementary Information and Supplementary Fig. 3). Therefore, the β -CASP domain seems to be a new example of the NBF superfold but is unlikely to bind nucleotides. This domain is highly conserved among CPSF-73 proteins, and probably has important functions in regulating their activity (see below).

The overall structure of yeast CPSF-100 (Fig. 1b) is remarkably similar to that of human CPSF-73, despite the low degree of sequence

conservation between them (see Supplementary Fig. 1b). The r.m.s. distance for 235 equivalent C α atoms in the metallo- β -lactamase domain (see Supplementary Fig. 2b) of the two structures is 2.2 Å (with 21% sequence identity), and that for 154 equivalent C α atoms in the β -CASP domain is 2.3 Å (with 17% identity).

There are, however, significant differences in the structures of CPSF-100 and CPSF-73 (see Supplementary Fig. 4). Most importantly, motifs for zinc binding in the metallo- β -lactamase domain are missing in CPSF-100 (see Supplementary Fig. 2b), and therefore this protein cannot bind zinc and is unlikely to possess nuclease activity. In addition, the β -CASP domain of CPSF-100 (Fig. 2b) is much larger than that of CPSF-73, but contains a highly flexible segment (see Supplementary Information and Supplementary Fig. 5).

Clear electron density for two zinc ions (confirmed on the basis of anomalous scattering data) was observed in the CPSF-73 active site (see Supplementary Fig. 6a), regardless of whether zinc was present in the crystallization solution. Even in the absence of added zinc, the zinc ions in the structure seem to have full occupancy as their temperature factor values are comparable to those of their protein ligands. This indicates that the active site of CPSF-73 might possess extremely high affinity for zinc ions, making it impossible to remove the zinc from the protein, consistent with the known resistance of 3' cleavage to relatively high concentrations of EDTA²². Two additional zinc ions were observed on the surface of CPSF-73 in the crystal grown in the presence of 0.5 mM zinc (see Supplementary Information and Supplementary Fig. 7).

The two zinc atoms in the active site are each bound in an octahedral environment, with a hydroxide ion as one of the bridging ligands (Fig. 3a, Supplementary Fig. 6b). The other bridging ligand is the side-chain carboxylate oxygen atom of Asp 179 from motif 4 (Fig. 3a). His 71, His 73 (motif 2) and His 158 (motif 3) provide three more ligands to the first zinc atom (Zn1), and Asp 75, His 76 (motif 2) and His 418 (motif C, see below) are the ligands to Zn2 (Fig. 3a). Finally, the two zinc ions are coordinated by oxygen atoms from a sulphate ion (Fig. 3a), which was present at 300 mM in the crystallization solution and has clear electron density (Supplementary Fig. 6a).

The structure of its active site strongly implies that CPSF-73 possesses hydrolase/nuclease activity. In canonical metallo- β -lactamases, the hydroxide that bridges the two zinc ions is the nucleophile for the hydrolysis reaction, and the substrate is directly liganded to the zinc atoms²³. In our structure, the hydroxide is directly below the sulphate group (Fig. 3a), which is probably a good mimic for the phosphate group of the pre-mRNA substrate at the cleavage site. Therefore, the hydroxide is located at the perfect position for an inline nucleophilic attack on the phosphate group to initiate the nuclease reaction (Fig. 3a).

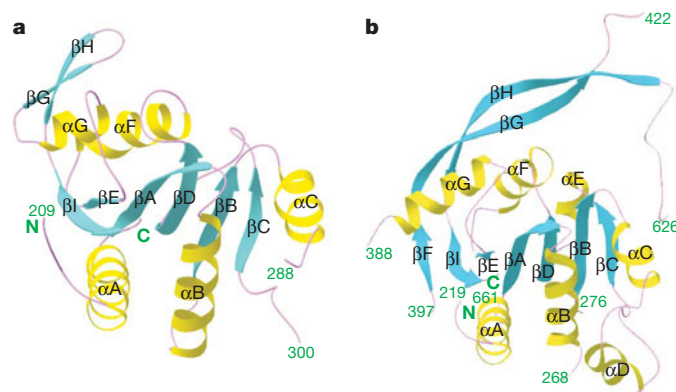


Figure 2 | The β -CASP domain of CPSF-73 and CPSF-100. Schematic drawings of the β -CASP domains of human CPSF-73 (a) and yeast CPSF-100 (b).

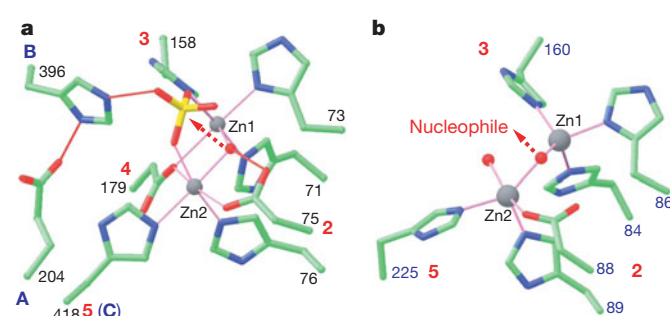


Figure 3 | The active site of CPSF-73. a, The zinc binding site for human CPSF-73. The motifs are labelled, and the bridging hydroxide ion is shown as a red sphere. Liganding interactions are indicated by thin magenta lines, and hydrogen-bonding interactions by thin red lines. The arrow indicates the nucleophilic attack from the hydroxide ion. b, The zinc binding site in L1 metallo- β -lactamase¹⁷.

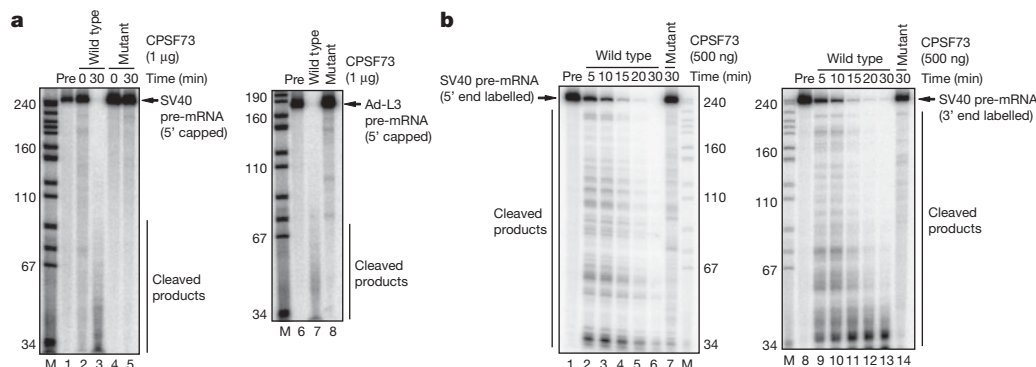


Figure 4 | CPSF-73 possesses endoribonuclease activity. **a**, Cleavage of 5'-capped SV40 late pre-mRNA (SVL) was performed with wild-type (lane 3) or mutant (lane 5) CPSF-73 at 30 °C for 30 min. Incubations for 0 min are shown in lanes 2 and 4. Cleavage of 5'-capped adenovirus L3 pre-mRNA (Ad-L3) is shown in lanes 7 and 8. Precursor RNAs (Pre) are shown in lanes 1

and 6. **M**, *MspI*-digested pBR322DNA markers. **b**, Time-course analysis of RNA cleavage by CPSF-73. Cleavage of 5' (lanes 2–6) and 3' (lanes 9–13) end-labelled SVL pre-mRNAs was performed with wild-type CPSF-73 (500 ng); time points were 5, 10, 15, 20 and 30 min. Mutant CPSF-73 (500 ng) did not exhibit RNase activity (lanes 7 and 14).

Our analysis indicates that His 396 (motif B, in the linker between the β -CASP domain and the metallo- β -lactamase domain; see Supplementary Fig. 2a) is the general acid for catalysis. This residue is hydrogen-bonded to the side chain of Glu 204 (motif A, last residue of strand β 12 in the metallo- β -lactamase domain, see Supplementary Fig. 2a) as well as an oxygen atom of the sulphate (Fig. 3a). This oxygen is located opposite to the direction of attack of the hydroxide (Fig. 3a), and therefore probably represents the leaving group in the hydrolysis reaction. After hydrolysis, the His 396 side chain can function as the general acid to protonate the oxyanion in the reaction product. This pair of Glu-His residues is also found in the structure of RNase Z, hydrogen-bonded to a phosphate group in the active site¹⁹. In yeast, mutation of either of these residues is lethal¹³.

Although an earlier sequence analysis indicated that motif B might be the best candidate for the zinc ligand in CPSF-73 (ref. 12), our structure shows that the ligand is the His residue from motif C. This residue is conserved among CPSF-73 proteins and its close homologues (RC-68 or Int11)^{24,25}, but not in the DNA nuclease Artemis^{12,26} (see Supplementary Fig. 1b). Therefore, the Zn2 atom might have a different coordination sphere in Artemis. The structure-based sequence alignment also indicates a change in the assignment of residues equivalent to motifs B and C in CPSF-100 (ref. 12; see Supplementary Fig. 1b), although neither residue is His in CPSF-100.

The binding modes of the zinc ions in CPSF-73 (Fig. 3a) are similar to that in RNase Z but distinct from those in canonical metallo- β -lactamases (Fig. 3b; see Supplementary Information). Another important difference between CPSF-73 and other metallo- β -lactamases is the presence of the β -CASP domain. Although most metallo- β -lactamases (Supplementary Fig. 2c), and RNase Z (Supplementary Fig. 2d), have an open zinc-binding site, the active site in CPSF-73 is located deep in the interface between the metallo- β -lactamase and the β -CASP domains (Fig. 1a). In fact, the β -CASP domain severely restricts access to the active site (Supplementary Fig. 8), and the scissile phosphate group probably cannot reach the zinc ions in the current structure. This indicates that a conformational change in the enzyme might be required for pre-mRNA binding (see Supplementary Information).

To show that CPSF-73 has endonuclease activity *in vitro* and to provide direct evidence for its role in 3'-end processing, we assessed the nuclease activity of the bacterially expressed CPSF-73, together with a derivative containing a double mutation (D75K/H76A) that destroys two of the zinc ligands in motif 2. We used a 5'-capped, uniformly labelled RNA containing the SV40 late polyadenylation site (SVL) as the substrate^{13,27}. Using standard conditions for 3' processing reactions, there was evidence for weak cleavage with the wild-type but not the mutant derivative (Supplementary Fig. 10a); a similarly prepared CPSF-100 derivative was also inactive (results not

shown). This induced us to attempt to optimize conditions to obtain more efficient cleavage.

Among the possibilities tested to increase nuclease activity was preincubation with divalent cations. Our structural studies showed that there are binding sites for additional Zn²⁺ on the surface (see Supplementary Fig. 7), and it is possible that (non-specific) binding of divalent cations to the protein surface could affect its activity. Preincubation with Zn²⁺ and all other divalent cations tested, except Ca²⁺, resulted in precipitation of the enzyme. Strikingly, after preincubation with Ca²⁺, the wild-type enzyme degraded the SVL substrate almost entirely to oligonucleotides, whereas the mutant was still inactive (Fig. 4a). A similar-sized RNA from adenovirus was also entirely degraded by wild-type but not mutant CPSF-73 (Fig. 4a). Preincubation with Ca²⁺, although required for enzyme activation (results not shown), was unnecessary for catalysis. It was diluted to 50 μ M in the reactions and nuclease activity was not sensitive to the addition of 5 mM EGTA (results not shown). Furthermore, there is no evidence that Ca²⁺ is required in authentic 3' processing, indicating that its activating function is probably performed by other components of the polyadenylation machinery. At lower CPSF-73 enzyme levels, partially degraded, higher-molecular-weight intermediates were seen (see Supplementary Fig. 10b). Significant cleavage required relatively high concentrations (10 ng μ l⁻¹ or ~200 nM), which we attribute to only a small fraction of the enzyme being activated by the preincubation.

Although our data are consistent with purified CPSF-73 possessing non-specific endonuclease activity, they do not conclusively rule out that it might have exonuclease activity, as previously suggested¹⁴. Indeed, there is evidence that Artemis has exonuclease activity²⁸. Although the presence of a 5' cap on the RNAs analysed above largely excludes 5'→3' exonuclease activity, we nonetheless compared the ability of CPSF-73 to degrade 5' and 3' end-labelled RNA (Fig. 4b). The two RNAs were degraded with comparable kinetics and generated similar patterns of intermediates, consistent only with endonuclease activity.

These results, together with previous data^{13,14}, establish CPSF-73 as a strong, sequence-independent endonuclease that functions with specificity in 3'-end processing of both polyadenylated and histone pre-mRNAs by interacting with other factors. In addition, our data indicate that the close sequence homologue of CPSF-73, RC-68 or Int11, is probably the endonuclease in the 3'-end processing of small nuclear RNAs²⁴.

METHODS

Protein expression, purification and crystallization. Residues 1–460 of human CPSF-73 were sub-cloned into the pET28a vector (Novagen) and overexpressed in *Escherichia coli* at 20 °C. The soluble protein was purified by nickel-agarose

affinity chromatography and gel-filtration chromatography. Yeast CPSF-100 (Ydh1, residues 1–720) was expressed and purified by nickel–agarose affinity chromatography, anion exchange and gel-filtration chromatography.

Crystals of human CPSF-73 were obtained at room temperature by the sitting-drop vapour diffusion method. In an attempt to increase the occupancy of zinc ions, some crystals were grown from a reservoir solution that also contained 0.5 mM ZnCl₂. Crystals of yeast CPSF-100 were obtained at 4 °C by the same protocol. A solution that had been infected by a fungus was crucial for the crystallization of this protein¹⁶. The crystals were frozen in liquid propane for data collection at 100 K.

Data collection, structure determination and refinement. X-ray diffraction data were collected on an ADSC CCD at the X4A beamline of Brookhaven National Laboratory. The diffraction data were collected at the zinc absorption peak for crystals of CPSF-73, and at the gold absorption peak for a KAu(CN)₂ derivative of CPSF-100. The structure of CPSF-73 was determined by the single-wavelength anomalous diffraction method²⁹, using the anomalous signal of zinc. The structure of CPSF-100 was determined by the single-isomorphous replacement method, supplemented with anomalous diffraction. The crystallographic information are summarized in Supplemental Table 2. Figs 1–3 were produced with Ribbons³⁰.

Pre-mRNA 3'-end cleavage assay. CPSF-73 was pre-incubated with 5 mM CaCl₂ at 37 °C for 30 min. Cleavage assays were carried out in reaction mixture (10 µl) containing ~1 ng labelled RNA substrates, 10 mM HEPES (pH 7.9), 10% (v/v) glycerol, 50 mM KCl, 0.25 mM DTT, 0.25 mM PMSF, 0.125 mM EDTA, 50 µM CaCl₂, RNase inhibitor (2 units), 500 ng BSA and indicated amounts of recombinant CPSF-73. Cleaved RNAs were isolated and fractionated on 6% urea PAGE. The data were analysed by Phospho-Imager.

Received 18 July; accepted 20 October 2006.

Published online 26 November 2006.

- Colgan, D. F. & Manley, J. L. Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* **11**, 2755–2766 (1997).
- Wahle, E. & Rueggsegger, U. 3'-end processing of pre-mRNA in eukaryotes. *FEMS Microbiol. Rev.* **23**, 277–295 (1999).
- Zhao, J., Hyman, L. & Moore, C. L. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* **63**, 405–445 (1999).
- Proudfoot, N. J. New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr. Opin. Cell Biol.* **16**, 272–278 (2004).
- Zorio, D. A. R. & Bentley, D. The link between mRNA processing and transcription: communication works both ways. *Exp. Cell Res.* **296**, 91–97 (2004).
- Wilusz, C. J., Wormington, M. & Peltz, S. W. The cap-to-tail guide to mRNA turnover. *Nature Rev. Mol. Cell Biol.* **2**, 237–246 (2001).
- Calvo, O. & Manley, J. L. Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev.* **17**, 1321–1327 (2003).
- Vinciguerra, P. & Stutz, F. mRNA export: an assembly line from genes to nuclear pores. *Curr. Opin. Cell Biol.* **16**, 285–292 (2004).
- Shatkin, A. J. & Manley, J. L. The ends of the affair: capping and polyadenylation. *Nature Struct. Biol.* **7**, 838–842 (2000).
- Wickens, M. & Gonzalez, T. N. Knives, accomplices, and RNA. *Science* **306**, 1299–1300 (2004).
- Daiyasu, H., Osaka, K., Ishino, Y. & Toh, H. Expansion of the zinc metallo-hydrolase family of the β -lactamase fold. *FEBS Lett.* **503**, 1–6 (2001).
- Callebaut, I., Moshous, D., Mornon, J.-P. & de Villartay, J.-P. Metallo- β -lactamase fold within nucleic acids processing enzymes: the β -CASP family. *Nucleic Acids Res.* **30**, 3592–3601 (2002).
- Ryan, K., Calvo, O. & Manley, J. L. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA* **10**, 565–573 (2004).
- Dominski, Z., Yang, X.-C. & Marzluff, W. F. The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. *Cell* **123**, 37–48 (2005).
- Kolev, N. G. & Steitz, J. A. Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs. *Genes Dev.* **19**, 2583–2592 (2005).
- Mandel, C. R., Gebauer, D., Zhang, H. & Tong, L. A serendipitous discovery that *in situ* proteolysis is required for the crystallization of yeast CPSF-100 (Ydh1p). *Acta Crystallogr. F62*, 1041–1045 (2006).
- Ullah, J. H. *et al.* The crystal structure of the L1 metallo- β -lactamase from *Stenotrophomonas maltophilia* at 1.7 Å resolution. *J. Mol. Biol.* **284**, 125–136 (1998).
- Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
- de la Sierra-Gallay, I. L., Pellegrini, O. & Condon, C. Structural basis for substrate binding, cleavage and allostery in the tRNA maturase RNase Z. *Nature* **433**, 657–661 (2005).
- Ishii, R. *et al.* Crystal structure of the tRNA 3' processing endoribonuclease tRNase Z from *Thermotoga maritima*. *J. Biol. Chem.* **280**, 14138–14144 (2005).
- Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the α - and β -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945–951 (1982).
- Hirose, Y. & Manley, J. L. Creatine phosphate, not ATP, is required for 3' end cleavage of mammalian pre-mRNA *in vitro*. *J. Biol. Chem.* **272**, 29636–29642 (1997).
- Garrity, J. D., Bennett, B. & Crowder, M. W. Direct evidence that the reaction intermediate of metallo- β -lactamase L1 is metal bound. *Biochemistry* **44**, 1078–1087 (2005).
- Baillat, D. *et al.* Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265–276 (2005).
- Dominski, Z., Yang, X.-C., Purdy, M., Wagner, E. J. & Marzluff, W. F. A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. *Mol. Cell. Biol.* **25**, 1489–1500 (2005).
- Moshous, D. *et al.* Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* **105**, 177–186 (2001).
- Takagaki, Y., Ryner, L. C. & Manley, J. L. Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. *Cell* **52**, 731–742 (1988).
- Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M. R. Hairpin opening and overhand processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781–794 (2002).
- Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58 (1991).
- Carson, M. Ribbon models of macromolecules. *J. Mol. Graph.* **5**, 103–106 (1987).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Ryan for discussions; B. Tweel for characterizing the fungus; R. Abramowitz, J. Schwanof and X. Yang for setting up the X4A beamline at the NSLS; and J. Khan and Y. Shen for help with data collection at the synchrotron. This research is supported in part by grants from the National Institutes of Health.

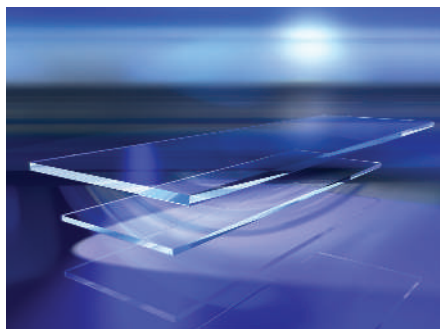
Author Information The atomic coordinates have been deposited at the Protein Data Bank (accession numbers 217T, 217V and 217X). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.T. (ltong@columbia.edu).

Growing pains

Protein microarrays are coming of age, and the development of specialized technologies is extending their high-throughput capabilities. Michael Eisenstein reports.

Like the younger child who has had an older sibling to 'soften up' his or her parents and make life a little easier, protein microarrays have benefited from lessons learned during the noisy adolescence of DNA microarrays. "Intellectually, the assays are identical to DNA microarrays," says microarray pioneer Mark Schena, visiting scholar at TeleChem International in Sunnyvale, California. "The basic tenets of miniaturization, automation and parallelism all hold for DNA and protein." Indeed, some tools of the trade, such as array spotters and readers, have needed virtually no changes to make the move to proteins. But the differences are there nonetheless, and the protein-microarray field is rapidly developing its own identity, with new challenges that require a diverse set of specialized tools and tricks.

One key distinction emerges early in the analysis process: the selection of array substrate. Unlike oligonucleotides, proteins are broadly heterogeneous in size, shape and chemistry, and the diversity of applications for protein arrays means that users need to shop around for the appropriate substrate. "We usually test about three or four surfaces, and see which one



Schott's Slide H is coated with a hydrogel that is a three-dimensional binding substrate for proteins.

gives us the best signal-to-noise ratio for a given assay," says Michael Snyder of Yale University.

There has been an explosion of options in the market, and several vendors — such as Xenopore of Hawthorne, New Jersey, and Schott of Jena, Germany — specialize almost exclusively in slides and substrates. Functionalized glass surfaces are a strong choice, as they virtually eliminate background fluorescence. According to Schena, TeleChem's SuperEpoxy glass slides are among the company's most popular prod-

ucts for protein-array construction. "The key benefit of the epoxide is that it's highly reactive at physiological conditions," he says. "Just depositing proteins on an epoxide-coated surface results in covalent linkage to the glass."

Three-dimensional substrates are also popular. For example, the Nexterion Slide H from Schott is coated with a covalently linked three-dimensional hydrogel with reactive groups that readily bind to proteins or peptides. "Slide H preserves the three-dimensional structure of proteins, providing a hydrophilic cell-like or cytosol-like environment, thereby maintaining stability and functionality," says Rüdiger Dietrich, Schott's director of research and development and technical support. Pam-Gené of s-Hertogenbosch in the Netherlands also takes advantage of a three-dimensional environment for its flow-through array platform. In its set-up, samples are pumped back and forth through a porous inorganic substrate where capture probes are immobilized.

Nitrocellulose, a timeless classic for protein work, remains among the most popular substrates. "We still don't have a substratum that's better than nitrocellulose for its binding

LOSING THE LABEL

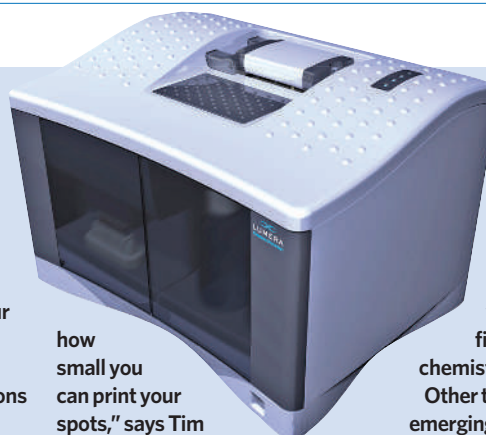
The appeal of using arrays free from fancy adornments — fluorescent, radioactive or otherwise — is fairly obvious. Such arrays could eliminate extra work and reduce errors in detection and analysis.

Surface plasmon resonance (SPR) is a promising technology for this approach. In SPR the interactants are fixed to a gold-coated substrate, and sample binding is detected as mass concentration-dependent changes in refractive index at that spot, which makes it possible to monitor binding in real time. "You can look at specificity, affinity, kinetics, make concentration measurements, and work in a range of different sample environments," says Gary Franklin, industrial-sector specialist at Biacore.

Recently acquired by GE Healthcare in Little Chalfont, UK, Biacore has pioneered the development of SPR platforms for a wide variety of proteomics applications. It has two main

options for array users. In the Flexchip platform, 400 interactants can be spotted on a slide and then screened against a single sample. In contrast, the A100 is limited to 20 immobilized interactants but, thanks to parallel flow systems, it can perform up to four simultaneous screens with large numbers of samples. "The A100 can look at up to 3,800 interactions per day in a variety of different, multiplexed ways," says Franklin.

Lumera of Bothell, Washington, is a relative newcomer to the market. It takes advantage of extremely rapid optical-switching technology, originally developed for telecommunications, in its ProteomicProcessor SPR instrument. Throughput was a key limitation of early SPR array experiments, and Lumera says its switching technology has reduced the problem. "We're basically limited by the size of the slide and



how small you can print your spots," says Tim Londergan, director of the company's bioscience business unit. "You can really see this tracking to DNA microarray scale." Lumera has nearly finished testing its 'beta' instrument and plans to launch its first commercial system in January 2007.

But SPR is not without limitations, and reduced sensitivity remains a common complaint. "My feeling is that there may be a somewhat limited dynamic range," says Joshua LaBaer, director of the

The ProteomicProcessor from Lumera will offer high-throughput label-free analysis.

Harvard Institute of Proteomics. "It's a fairly narrow window, and you have to be able to figure out how to get your chemistry within that window."

Other technologies are also emerging, such as the arrays based on microcantilevers made by Protiveris of Rockville, Maryland, or the atomic-force microscopy system from BioForce Nanosciences of Ames, Iowa. Most of these platforms are in their infancy, but technology development is under way and many microarray users foresee a label-free future — one way or another. "It's going to revolutionize the way we think about things," says LaBaer. "It won't happen in ten days or a year, but it's going to happen."

M.E.

capacity — it's cheap and can be mass-produced," says Emanuel Petricoin, co-director with Lance Liotta of the Center for Applied Proteomics and Molecular Medicine at George Mason University in Manassas, Virginia. "The problem is that it has high background fluorescence and you have to come up with different labelling strategies to get around that." One solution uses an ultrathin nitrocellulose layer — such as the PATH slides from GenTel BioSciences of Madison, Wisconsin — which maintains binding capacity but reduces background noise.

More specialized surfaces enable orientation-specific presentation of protein probes, such as nickel-NTA (nitrilotriacetic acid) surfaces for use with polyhistidine-tagged proteins. Lumera in Bothell, Washington, uses a protein-protein interaction-based immobilization format for its arrays. "We have a proprietary protein-tag technology that's composed of two coiled coils that bind together with picomolar affinity," says product group manager Ronald Dudek. "One can be engineered into a protein or antibody library, the other is part of the surface chemistry." The pairing of streptavidin-coated slides with biotinylated proteins is also widely used.

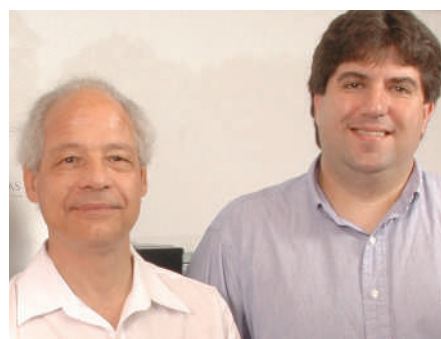
Seeing the light

One aspect of arrays that has changed relatively little is the dominance of fluorescence as a method of detection. "People generally use the standard Cy3 and Cy5 fluorophores, which are cheap, reproducible, and scanners are already configured to read them," says Petricoin. His group has done initial studies with

semiconductor quantum dots, which offer the benefits of robust long-term fluorescence and enhanced potential for multiplexing. But for many investigators, generically reactive fluorescent dyes are sufficient. "In some ways, labelling is easier than with DNA," says Schena, "because proteins are naturally highly reactive with fluorescent reagents."

Some manufacturers improve the quality of fluorescent-array experiments with specialized instrumentation. The ZeptoREADER from Zeptosens in Witterswil, Switzerland, uses 'planar waveguide' technology, in which an evanescent electromagnetic field is generated by directing light into a specially designed array substrate through a diffraction grating. "The key benefit is that you excite only fluorophores that are bound on the surface," says managing director Markus Ehrat. "In many conventional assays, you don't need to separate bound from unbound fluorophores." Alternatively, in PamGene's flow-through system, samples are repeatedly cycled through porous arrays while being imaged by a sensitive charge-coupled device camera, making it possible to take real-time kinetic measurements.

For experiments involving particularly small samples or scarce protein targets, it might be necessary to amplify the signal. One way to do this is rolling-circle amplification (RCA), first adapted for use with protein arrays by David Ward at Yale University. In RCA, circular DNA molecules are hybridized to capture probes that are conjugated to a detection antibody. These are then repeatedly replicated by DNA



Lance Liotta (left) and Emanuel Petricoin use reverse-phase arrays instead of sandwich arrays in their clinical research.

polymerase and detected with complementary fluorescent probes. Another alternative involves enhanced chemiluminescent detection with tyramide signal amplification, a system available from PerkinElmer of Wellesley, Massachusetts.

As effective as these techniques have proven, they all require a degree of tagging or modification, which can be time-consuming or even impractical for certain samples, such as clinically derived preparations. So many researchers are keeping their eyes on the evolution of 'label-free' technologies that could greatly simplify future studies (see 'Losing the label').

As with genomic arrays, the earliest application — and still the main use — of protein arrays has been the detection or quantification of targets from a research sample or clinical preparation. This typically involves a 'capture array'

AN APT SOLUTION?

Larry Gold has a solution for the problems faced by users of antibody arrays — stop using antibodies. Gold, founder and chief executive of SomaLogic in Boulder, Colorado, is outspoken in his belief that aptamers — small nucleic-acid molecules with specialized functional properties — should replace antibodies in capture arrays.

Years ago, Gold and his team tried to calculate how much multiplexing could be done with an antibody array before background noise became a significant issue. "We guessed that sandwiched antibody arrays were going to start running into serious noise problems when there were 20 analytes per array," Gold says. He thinks that aptamers could provide a level of specificity that surpasses what most antibodies achieve.

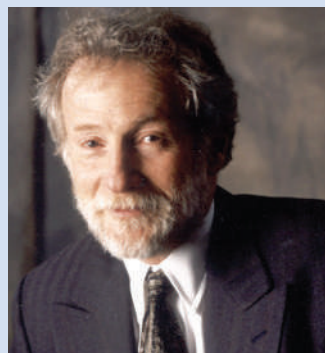
More than 15 years ago, Gold was one of the inventors of a now widely used procedure called SELEX — systematic evolution of

ligands by exponential enrichment. In this system, multiple rounds of selection and amplification can be used to select for DNA or RNA molecules with high specificity for a target of choice.

Gold and his colleagues have since enhanced the procedure, and SomaLogic uses a high-throughput version of SELEX to generate 'photoaptamers', which can be covalently crosslinked to bound targets following irradiation. As a result, aptamers with high affinity and specific crosslinking can be used to measure proteins in complex samples without needing the extra specificity provided by secondary antibodies.

Another advantage of the aptamer platform is simplicity of detection — once protein targets are bound, they can be labelled with a generic protein-binding fluorescent dye, eliminating the need for analyte-specific sandwich detection reagents.

Early platform tests have been promising — SomaLogic is achieving success rates of about 80%, Gold says, and he believes a product launch is imminent. "With the right partner, we could launch a product for the research market within some months," he says, "but we think diagnostics is key, and we're even more interested in transforming evidence-based healthcare."



Larry Gold believes his aptamer arrays will succeed in applications where antibody arrays fall short.

Some antibody experts are impressed by aptamers, but doubt whether they will unseat the current king. "What Larry Gold and others have shown is that when it works, it works very, very well," says Mathias Uhlén of the Royal Institute of Technology in Stockholm, Sweden, "although I would be surprised if aptamers are the dominant scaffold in the future, due to limitations in the chemical space."

But SomaLogic is banking on the use of novel nucleotides to increase this chemical space, and believes that a full proteome complement of aptamers is unnecessary — so that high specificity against a few thousand well-chosen targets could be more than sufficient. "We think that, for diagnostics, there's so much redundancy in biology that you'll be able to do useful biomarker discovery with an incomplete proteome that's still quite large," says Gold.

M.E.

in which selected antibodies are spotted on to a substrate and are then exposed to the sample. Detection can be performed directly with a labelled sample, although this can increase the background noise. Many researchers prefer 'sandwich' formats in which a second, tagged antibody is applied that recognizes a different epitope from the capture antibody. This considerably improves sensitivity and the signal-to-noise ratio, but also requires two high-quality antibodies for every target.

Antibody arrays have been commercially available for some time, but companies have taken different approaches to them. Sigma-Aldrich of St Louis, Missouri, for example, has several arrays in its Panorama product family, each containing 100–200 antibodies targeting proteins involved in processes such as signal transduction or gene regulation. The company also plans to release a broad-content array with more than 700 antibodies. "Our current pathway arrays will be included, in addition to various other signalling and regulatory gene-product antibodies," says market-segment manager Richard Pembrey.

TeleChem, on the other hand, leaves antibody selection to the users for its ArrayIt microarrays. "For now, it's custom work just because a full proteomic complement of antibodies is still expensive for most researchers," says Schena. "But our printing technology scales pretty nicely, and our products can span a pretty wide spectrum, between 100 elements and 50,000."

After encountering limited success with capture arrays in their clinical research, Liotta

and Petricoin developed the reverse-phase protein array, in which samples are spotted on the array and then probed with detection antibodies, requiring only one antibody per analyte. Much of their work has centred on signalling pathways in human disease, and this technique has worked well with their research, they say. "From a few thousand cells obtained by laser microdissection, we can look at hundreds of phosphorylation endpoints quantitatively, and look at a target and all the downstream signalling around it," says Liotta.

Such sample-specific arrays are difficult to commercialize, although Zeptosens is attempting to address the needs of this community with its cell lysate array (CeLyA) product line, which provides users with protocols, reagents and equipment to prepare chips for reverse-phase experiments. The company also has an active service division. "We deposit roughly the content of one or two cells per spot," says Ehrat, "and we can monitor changes of 15% from control to a treated sample."

But both approaches face a key limitation: antibodies. "A lot of users have two problems," says Mathias Uhlén, a researcher at the Royal Institute of Technology in Stockholm, Sweden, and chair of the Human Antibody Initiative of the Human Proteome Organisation (HUPO). "One is that you buy reagents, and half of them don't work in your application, and the second is that it's not easy to buy 200 antibodies that work on a single platform." Part of the solution lies in thorough validation, but other problems arise from the feature density of today's arrays

and the broad range of protein expression. "Even if you have an antibody with picomolar affinity for your target, where the background is micromolar; if you have cross-reactivity against a protein that is 10^6 times more abundant, you will see that protein first," Uhlén says. Some are exploring alternative affinity reagents, such as recombinant single-chain antibodies or nucleic-acid aptamers (see 'An apt solution?'), but most in the field still see the limits of antibodies as secondary to their strengths. "So far, the good old antibody is still going strong," says Uhlén.

All together now

High-content protein microarrays have brought the classic protein–protein interaction assay to levels of throughput previously only possible with two-hybrid assays. These 'proteome chips' originated in Snyder's lab, in the form of arrays composed of protein products from nearly 6,000 yeast open reading frames. These chips are now available from Invitrogen of Carlsbad, California, which has continued to develop these and other arrays as part of its ProtoArray product line. Invitrogen sells both human and yeast proteomic arrays in a variety of formats, designed for use in protein–protein interaction assays as well as functional studies.

Developing arrays of soluble — and ideally, functional — protein at this scale poses considerable challenges, says Paul Predki, vice-president of proteomics at Invitrogen. "Our latest array product has more than 8,000 human proteins, and you can imagine the challenges

(ALMOST) NO ASSEMBLY REQUIRED

Each stage in the building of a protein chip — expression, purification, immobilization — adds a layer of experimental complexity, as each feature may need its own optimization process to ensure consistent quality. "We thought we needed a better way to do this," says Joshua LaBaer, director of the Harvard Institute of Proteomics.

The solution that he and his team arrived at was the 'nucleic acid programmable protein array' (NAPPA), in which cDNAs encoding GST fusion proteins are arrayed on chips alongside antibodies that recognize GST. The array is then subjected to cell-free transcription and translation; as protein is produced, it gets bound by an antibody and presented for analysis. According to LaBaer, NAPPA has simplified his group's research. "You don't have to purify proteins — you just purify DNA, so it's pretty easy, and it's been successful for printing about 95%



Joshua LaBaer uses self-assembling arrays to bypass problems with proteins.

to 96% of the things we make," he says. "And when we do protein–protein interactions, we're getting interactions that make sense and not a lot of false positives."

Although initial arrays were limited in size, LaBaer and his team have since generated NAPPA arrays with up to 2,000 features, and they hope to surpass this soon.

Other techniques even bypass DNA immobilization. In the protein

in situ array (PISA) developed by Michael Taussig of the Babraham Institute in Cambridge, UK, cDNAs are amplified *in situ* with primers that encode polyhistidine tags, so that proteins can be captured on a nickel-NTA-coated surface.

More recently, Philipp Angenendt of the German Cancer Research Center in Heidelberg transferred the PISA principle to a microarray set-up, integrating cell-free

production of histidine-tagged protein from unpurified PCR fragments with a multiple spotting technique (MIST) previously developed by his group. MIST uses automation to apply array reagents precisely and sequentially to specific spots. As a result, each transcription/translation reaction is confined to a tiny, sub-nanolitre droplet, allowing greater density — up to 13,000 spots at present. "The nice thing about it is that the proteins expressed remain in a liquid environment," says Angenendt, "and the structure should be as intact as it can be with a solid-phase immunoassay format."

Both Angenendt and LaBaer are now fine-tuning their processes. "We're really working on large-scale screens, doing biomarker and protein interaction studies," says LaBaer. "We've also got some preliminary enzymatic data that look promising."

M.E.



JPT uses a high-throughput, non-contact printing process to assemble its peptide arrays.

of doing quality control for the functionality of all of these proteins, especially when for some 15% we don't even know the function, and other proteins have multiple functions," he says. Invitrogen bypasses some of the risks of misfolding or improper processing by expressing proteins in insect cells rather than in bacteria. "Almost all of our human proteins are expressed using baculovirus and we've found, for example, much higher success rates in obtaining active kinases," says Predki.

Snyder is also continuing his work with the yeast arrays, which now cover roughly 75–80% of the yeast proteome. These proteins have been cloned as both C- and N-terminal glutathione S-transferase (GST) fusions to improve the odds of obtaining well-folded fusions, although some proteins refuse to comply. Several groups are now attempting to resolve this issue with chip formats in which transcription and translation of cloned genes are performed on the chip (see 'Almost no assembly required').

Proteome chips also have clinical promise, and Snyder suggests that they could be valuable for drug screening. "Any drug company with a set of lead compounds should put them on a protein chip and see what the possible side effects might be, or what their targets are," he says. "I think that will be the future."

Pathogen-derived proteome chips may also have clinical value — Snyder recently developed a coronavirus protein array for diagnostic use, and Joshua LaBaer, director of the Harvard Institute of Proteomics in Boston, Massachusetts, has developed full proteomic arrays for the pathogens *Francisella tularensis* and *Vibrio cholerae*. "The idea would be to take serum from individuals who have been infected, both pre- and post-immune response, and look at

the proteins against which the patient has mounted a response," says LaBaer.

JPT Peptide Technologies in Berlin, Germany, is examining similar applications, with microarrays displaying overlapping peptides comprising the full *Mycobacterium tuberculosis* proteome. "We started with a small subset of 10,000 peptides from tuberculosis markers," says managing director Michael Schutkowski, "and found that the chip is more predictive and sensitive than all the other tuberculosis tests on the market."

Where form meets function

Functional information is available for only a subset of the more than 23,000 proteins in the human proteome, and much of this is incomplete or even anecdotal. A truly comprehensive understanding of biological processes requires the building of functional proteomic maps, and protein arrays could be a perfect tool for this.

Snyder and his team made a step forward when they applied their proteome chips to the assembly of a yeast phosphorylation map, incubating their microarrays with 87 different yeast protein kinases, and using radiolabelled ATP to identify proteins that were being modified. "We showed that closely related kinases had different specificities," says Snyder. His lab was also able to combine its findings with other genomic and proteomic data to identify regulatory 'modules' of interacting proteins that seem to be conserved in other eukaryotic species. "In my mind, that's the power of '-omics,'" he says. "We can come up with new principles by looking at these large data sets."

Invitrogen's arrays are also suitable for such assays. "The same application is fully validated on our human array," says Predki. And Sigma-Aldrich has a product for kinomic analysis, the Panorama Human Kinase v1 array, based on technology licensed from Procognia in Maidenhead, Berkshire, UK. This array incorporates 152 different human kinases for the identification of interacting partners and substrates, or the analysis of putative therapeutic compounds. On the clinical side, Liotta and Petricoin have found success using phosphospecific antibodies with reverse-phase arrays to characterize differential phosphorylation in biopsy specimens, and their arrays are now undergoing clinical trials as a diagnostic or prognostic tool for cancer.

Several groups are also developing array formats for glycan analysis. Brian Haab of the Van Andel Institute in Grand Rapids, Michigan, has used lectins — a family of proteins with specific binding preferences for different types of glycans — as a probe for monitoring glycosylation patterns of array-bound proteins. "We now work with about 20 different lectins," says Haab, "and we

recently demonstrated that we can do partial digestion of glycan structures using glycosidases to expose additional underlying structures to lectins, allowing us to get more complete information." These are mainly being used to characterize glycosylation changes in disease states, and Haab has licensed his platform to GenTel for commercial development.

Snyder, meanwhile, is interested in using such arrays for discovery, and has used polyclonal antibodies against yeast glycans to reveal hundreds of previously unidentified glycoproteins. QIAGEN of Venlo, the Netherlands, will soon launch a glycomics array product, the QProteome GlycoArray kit, based on technology licensed from Procognia, which uses immobilized lectins as a capture reagent.

Peptide arrays are a useful alternative for functional assays, and JPT has applied a peptide-synthesis process to develop a variety of array products. "The peptide chip is much more stable compared with a protein chip," says Schutkowski, "and because we have the peptides beforehand in a microtitre plate, we can analyse each peptide for purity and whether it has the right content and the right modifications before we immobilize it." JPT has used literature- and data-mining to design peptide libraries containing thousands of putative kinase targets, which can be used to identify recognition sequences for tyrosine or serine/threonine kinases. The company also offers similar arrays for phosphatase and protease target identification, as well as 'random' arrays that can potentially reveal previously unidentified target sites.

PamGene uses a peptide-based format in its PamChip arrays for analysing the kinetics of tyrosine and serine/threonine kinases against hundreds of arrayed probes. "IC₅₀ results, selectivity data, mechanism of action information and multiple rate constants can all be obtained in one experimental run," says PamGene's vice-president of technology, Rinie van Beuningen.

As these various applications strive towards full maturity, those in the protein microarray field are wrestling with the same issues that have confronted users of DNA arrays — experimental standardization and accurate quantification. The former issue is now being investigated by HUPO, which is looking to develop guidelines for experimental design and data annotation. But the key to accurate quantification will most likely lie in technological evolution. "Right now these data are just

semi-quantitative, but I think that with the right technologies they could be made completely quantitative," says Snyder. "I'd like to see a future where people can just think up an experiment, buy an array, and do it."

Michael Eisenstein is the former technology editor for *Nature* and *Nature Methods*.



Michael Snyder is looking to a future when people can think up an experiment, buy an array, and do it.

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
Catalogue and custom protein arrays			
Alta Bioscience	Custom peptide-array printing services	Birmingham, UK	www.altabioscience.bham.ac.uk
BD Biosciences	Cytometric bead arrays for multiplex protein analysis in solution, catalogue antibodies	San Jose, California	www.bdbiosciences.com
Cambridge Peptides	Peptide synthesis and purification services	Cambridge, UK	www.cambridgepeptides.com
Clontech	Premade antibody microarrays	Mountain View, California	www.clontech.com
EMD Biosciences	Human and murine cytokine antibody arrays, various catalogue antibodies	Darmstadt, Germany	www.emdbiosciences.com
Epitome Biosystems	Custom-designed antibody arrays	Waltham, Massachusetts	www.epitomebiosystems.com
Eurogentec	Custom protein, peptide and antibody arrays, custom and catalogue antibodies	Seraing, Belgium	uk.eurogentec.com
Fluidigm	BioMark microfluidic dynamic array immunoassay platform	South San Francisco, California	www.fluidigm.com
Hypromatrix	AntibodyArrays and Staining AntibodyArrays, catalogue antibodies	Worcester, Massachusetts	www.hypromatrix.com
INTAVIS	Custom peptide sets and arrays, custom polyclonal antibodies	Cologne, Germany	www.intavis.com
Invitrogen	ProtoArray human and yeast proteome chips, BioSource Luminex assays	Carlsbad, California	www.invitrogen.com
JPT Peptide Technologies	Predesigned and custom peptide arrays for functional proteomics	Berlin, Germany	www.jpt.com
LC Sciences	Peptide arrays for phosphoproteomics, epitope mapping	Houston, Texas	www.lcsciences.com
New England Peptide	Custom peptide arrays	Gardner, Massachusetts	www.newenglandpeptide.com
PamGene	PamChip flow-through peptide microarrays for functional studies	's-Hertogenbosch, the Netherlands	www.pamgene.com
Panomics	Domain and transcription factor arrays, antibody and phosphoproteomic arrays	Fremont, California	www.panomics.com
Pepscan Systems	PepChip peptide arrays for enzyme profiling, epitope mapping	Lelystad, the Netherlands	www.pepscan.nl
Peptide Specialty Laboratories	Specialized peptide sets for various applications	Heidelberg, Germany	www.peptid.de
Procognia	Specialized arrays for kinomics and glycomics	Berkshire, UK	www.procognia.com
Protagen	UNChip protein arrays	Dortmund, Germany	www.protagen.de
ProteinOne	Active Protein Array platform for profiling transcription factors, nuclear receptors and cancer proteins	Bethesda, Maryland	www.proteinone.com
Protneomix	Protein and antibody fabrication and screening services	Nantes, France	www.protneomix.com
QIAGEN	Qproteome GlycoArray kit, Liquechip multiplex bead-based arrays	Venlo, the Netherlands	www.qiagen.com
RayBiotech	A variety of specialized antibody arrays	Norcross, Georgia	www.raybiotech.com
Sigma-Aldrich	Panorama antibody arrays, functional arrays and tissue extract arrays	St Louis, Missouri	www.sigmaaldrich.com
TeleChem International	Antibody arrays, microarraying instruments, slides and substrates, reagents	Sunnyvale, California	www.arrayit.com
Upstate (Millipore)	Beadlyte xMAP-based multiplex assays	Billerica, Massachusetts	www.upstate.com
US Biomax	Premade fluorescent or colorimetric antibody capture arrays	Ijamsville, Maryland	www.biomax.us
Whatman	FAST Macro antibody arrays, FAST slides and reagents, array services	Middlesex, UK	www.whatman.com
Zeptosens	CelyA reagents and equipment for reverse-phase array preparation, array services	Witterswil, Switzerland	www.zeptosens.com
Zyomyx	Human and murine cytokine profiling arrays	Hayward, California	www.zyomyx.com
Systems for array production			
ArrayJet	Inkjet-based microarray printers	Mayfield, Scotland	www.arrayjet.co.uk
BioAutomation	MagnaSpotter for pin-based microarray printing	Plano, Texas	www.bioautomation.com
Bio-Rad Laboratories	Arraying systems, hybridization chambers, software tools	Hercules, California	www.bio-rad.com
Genetix	Pin-based arrayers and array scanners	New Milton, UK	www.genetix.com
Genomic Solutions	OmniGrid microarrayers and hybridization chambers	Ann Arbor, Michigan	www.genomicsolutions.com
LabNEXT	Equipment for array printing and processing	Glenview, Illinois	www.labnext.com
PerkinElmer	Equipment for array production and scanning	Wellesley, Massachusetts	las.perkinelmer.com
Scienion	scifLEXARRAYER Piezo Dispenser for array manufacture, specialized array slide formats	Berlin, Germany	www.scienion.de
V&P Scientific	Glass slide and compact disc microarrayers	San Diego, California	www.vp-scientific.com
Array scanners			
Alpha Innotech	Array scanners	San Leandro, California	www.alphainnotech.com
Applied Precision	arrayWoRx biochip reader	Issaquah, Washington	www.api.com
Biomedical Photometrics	Microarray scanners and analytical software	Waterloo, Canada	www.confocal.com
Biomolex	Microarray readers based on isotopic detection	Oslo, Norway	www.biomolex.com
BMG LABTECH	Microplate readers for fluorescence and luminescence analysis	Offenburg, Germany	www.bmglabtech.com
Fujifilm	Scanners and analytical software	Tokyo, Japan	www.fujifilm.com
LaVision BioTec	BioAnalyzer CCD-based platforms for microarray imaging	Bielefeld, Germany	www.lavisionbiotec.de
MiraBio	FMBIO fluorescent scanner	Alameda, California	www.mirabio.com
Molecular Devices	Microarray scanners and analytical software	Sunnyvale, California	www.moleculardevices.com
Tecan	Microarray scanners and hybridization stations	Zurich, Switzerland	www.tecan.com
VIDAR Systems	Revolution 4200 microarray scanner	Herndon, Virginia	www.microarrayscanner.com
Antibodies and affinity reagents			
21st Century Biochemicals	Custom antibody production and peptide synthesis	Marlboro, Massachusetts	www.21stcenturybio.com
Abcam	Catalogue antibodies and protein A-derived affibodies	Cambridge, UK	www.abcam.com
Abgent	Antibodies and antibody production	San Diego, California	www.abgent.com
Affibody	Affibody non-immunoglobulin affinity molecules	Bromma, Sweden	www.affibody.com
AnaSpec	Antibodies and custom antibody production, peptide synthesis	San Jose, California	www.anaspec.com
Antibodies Incorporated	Antibodies and antibody production	Davis, California	www.antibodiesinc.com
BioDesign	Antibodies and monoclonal production	Saco, Maine	www.biodesign.com
BioGenes	Monoclonal and polyclonal antibodies, custom peptide synthesis	Berlin, Germany	www.biogenes.de
BioLegend	Antibodies and fluorescent dyes	San Diego, California	www.biolegend.com
Biomeda	Monoclonal and polyclonal antibodies, lectins	Foster City, California	www.biomeda.com
BioTrend	Monoclonal and polyclonal antibodies, custom peptides	Cologne, Germany	www.biotrend.com

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
Cell Sciences	Monoclonal and polyclonal antibodies	Canton, Massachusetts	www.cellsciences.com
Cell Signaling Technology	Monoclonal and polyclonal antibodies, screening kits and assays	Danvers, Massachusetts	www.cellsignal.com
Chemicon (Millipore)	Custom and catalogue antibodies	Billerica, Massachusetts	www.chemicon.com
KPL	Catalogue and custom antibodies, protein labelling reagents	Gaithersburg, Maryland	www.kpl.com
Lampire Biological Laboratories	Custom monoclonal and polyclonal antibody production	Pipersville, Pennsylvania	www.lampire.com
Leinco Technologies	Catalogue and custom antibodies	St Louis, Missouri	www.leinco.com
Open Biosystems	Custom and catalogue polyclonal and monoclonal antibodies	Huntsville, Alabama	www.openbiosystems.com
QED Bioscience	Catalogue and custom antibodies	San Diego, California	www.qedbio.com
Rockland Immunochemicals	Antibodies and custom antibody production, detection reagents	Gilbertsville, Pennsylvania	www.rockland-inc.com
Santa Cruz Biotech	Catalogue polyclonal and monoclonal antibodies	Santa Cruz, California	www.scbt.com
Southern Biotech	Custom and catalogue monoclonal and polyclonal antibodies	Birmingham, Alabama	www.southernbiotech.com
Stratagene	Selected catalogue antibodies	La Jolla, California	www.stratagene.com
US Biological	Custom and catalogue antibodies, lectins	Swampscott, Massachusetts	www.usbio.net
VWR	Antibodies from various suppliers, custom antibodies from NeoClone	West Chester, Pennsylvania	www.vwr.com

Slides and array substrates

Accelr8	OptArray microarraying slides	Denver, Colorado	www.accelr8.com
Erie Scientific	Glass slides and substrates for microarrays	Portsmouth, New Hampshire	www.eriemicroarray.com
Full Moon BioSystems	Protein microarray slides and services	Sunnyvale, California	www.fullmoonbio.com
GE Healthcare	CodeLink activated slides, labelling reagents	Little Chalfont, UK	www.amershambiosciences.com
GenTel BioSciences	Slides, substrates and reagents for protein microarrays, array services	Madison, Wisconsin	www.gentelbio.com
Greiner Bio One	High-throughput microarray plates	Monroe, North Carolina	www.greinerbioone.com
Microsurfaces	Surface coating for array slides	Minneapolis, Minnesota	www.proteinslides.com
Nunc Brand	Microarray slide coatings	Roskilde, Denmark	www.nuncbrand.com
Pierce Biotechnology	PATH protein microarray slides	Rockford, Illinois	www.piercenet.com
Schott North America	Slides, substrates and microarray reagents	Louisville, Kentucky	www.us.schott.com/nexterion
Xenopore	Coated microarray slides, arrayers, array scanning services	Hawthorne, New Jersey	www.xenopore.com

SPR and other platforms for label-free array analysis

Biacore (GE Healthcare)	SPR instruments for high-throughput interaction and kinetics analysis	Uppsala, Sweden	www.biacore.com
BioForce Nanosciences	Atomic-force-microscopy-based platform for nanoarray analysis	Ames, Iowa	www.bioforcenano.com
Bio-Rad Laboratories	ProteOn XPR36 SPR-based protein interaction array system, Bio-Plex suspension arrays	Hercules, California	www.bio-rad.com
Eco Chemie	Autolab SPR instrumentation and accessories	Utrecht, the Netherlands	www.ecochemie.nl
Genoptics	SPR platforms for performing array studies	Orsay Cedex, France	www.genoptics-spr.com
GWC Technologies	SPR platforms suitable for protein microarray studies	Madison, Wisconsin	www.gwctechnologies.com
IBIS Technologies	SPR products and sensors	Hengelo, the Netherlands	www.ibis-spr.nl
ICx Technologies	SensiQ SPR system for biomolecular interaction analysis	Washington DC	www.icxt.com
Lumera	Proteomic-Processor SPR instrument and array substrates	Bothell, Washington	www.lumera.com
NanoNord	Cantion's cantilever technology for interaction detection	Aalborg, Denmark	www.nanonord.com
Plasmonic	Products and instruments for SPR	Wallenfels, Germany	www.plasmonic.de
Protiveris	Micro- and nanocantilever arrays for protein interaction analysis	Rockville, Maryland	www.protiveris.com
SRU Biosystems	BIND Biosensor technology for label-free interaction analysis	Woburn, Massachusetts	www.srubiosystems.com
XanTec	SPR biosensors, and array-compatible slides	Münster, Germany	www.xantec.com

Other array products

Advantix	ArrayBooster for improved hybridization of protein arrays	Brunnthal, Germany	www.advantix.de
Ariadne Genomics	Pathway Studio software for interpreting microarray data	Rockville, Maryland	www.ariadnegenomics.com
Gyros	Gyrolab platform for protein quantification immunoassays	Uppsala, Sweden	www.gyros.com
Imaging Research	ArrayVision software for array analysis	Linton, UK	www.imagingresearch.com
Kreatech Biotechnology	ULS labelling technology for use with protein arrays	Amsterdam, the Netherlands	www.kreatech.com
Luminex	xMAP microsphere platform for high-throughput, solution-based multiplex assays	Austin, Texas	www.luminexcorp.com
Meso Scale Discovery	Electrochemiluminescence-based platform for array analysis	Gaithersburg, Maryland	www.meso-scale.com
Promega	HaloTag system for protein labelling	Madison, Wisconsin	www.promega.com
Roche Applied Science	Reagents for protein labelling	Indianapolis, Indiana	www.roche-applied-science.com
SomaLogic	Aptamer-based chips for protein detection and quantitation	Boulder, Colorado	www.somallogic.com
SurModics	Microarray stabilizer reagents	Eden Prairie, Minnesota	www.surmodics.com
VigeneTech	MicroVigene array analysis software	North Billerica, Massachusetts	www.vigenetech.com

● see advertisement

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

Sometimes, it is the smallest events that have the biggest impact. Seemingly trivial details can suddenly offer incredible insight, an off-the-cuff remark can provide the best advice you've ever had. So it would seem for the four graduate students who have kept monthly journals for *Naturejobs* this past year.

Mhairi Dupré, a first-year graduate student in evolutionary developmental biology at the University of Oxford, UK, nearly didn't do her PhD at all. Her first doctoral programme proved to be unsatisfying and she took some time out after halting her studies. She then decided to give academia another try and landed a place at Oxford. Now she finds her thoughts turning to experimental design in the middle of planning dinner — a sign that her return to the lab was the right choice.

For Andreas Andersson, a PhD in oceanography in Hawaii sounded like it would be an adventure. But he found that surfing the island's legendary waves provided more immediate gratification. Back at the bench, he wondered whether a scientific career would give him the same buzz. He now acknowledges that the thrill of data collection is worth some slightly more mundane work before and after each dive.

Molecular-biology graduate student Milan de Vries, who is at the Massachusetts Institute of Technology, openly admits that yeast isn't glamorous — even though he sometimes sings to his samples. But he realized, through the illness of a relative, that his work in basic science could have more profound implications and that the daily slog is worth pursuing.

And for Katja Bargum, a graduate student at the University of Helsinki, Finland, clarity came amid ending a relationship, buying a home and skiing. She realized that she enjoyed writing about research more than doing it, and is now pursuing a career in science journalism.

These four students are now closing their diaries, ready to pursue the next step in their careers. You can see their final journal entries, which look back at what they've learnt, online at www.nature.com/naturejobs/archive/archive-gradjn.html.

Paul Smaglik, *Naturejobs* editor

Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
Natureevents:
Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestranguez (4994)

UK/Ireland/Italy/RoW:
Loredana Milanese (4944)
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

European Satellite Office
Germany: Patrick Phelan
Tel: +49 89 54 90 57 11
Fax: +49 89 54 90 57 20
e-mail: p.phelan@nature.com

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

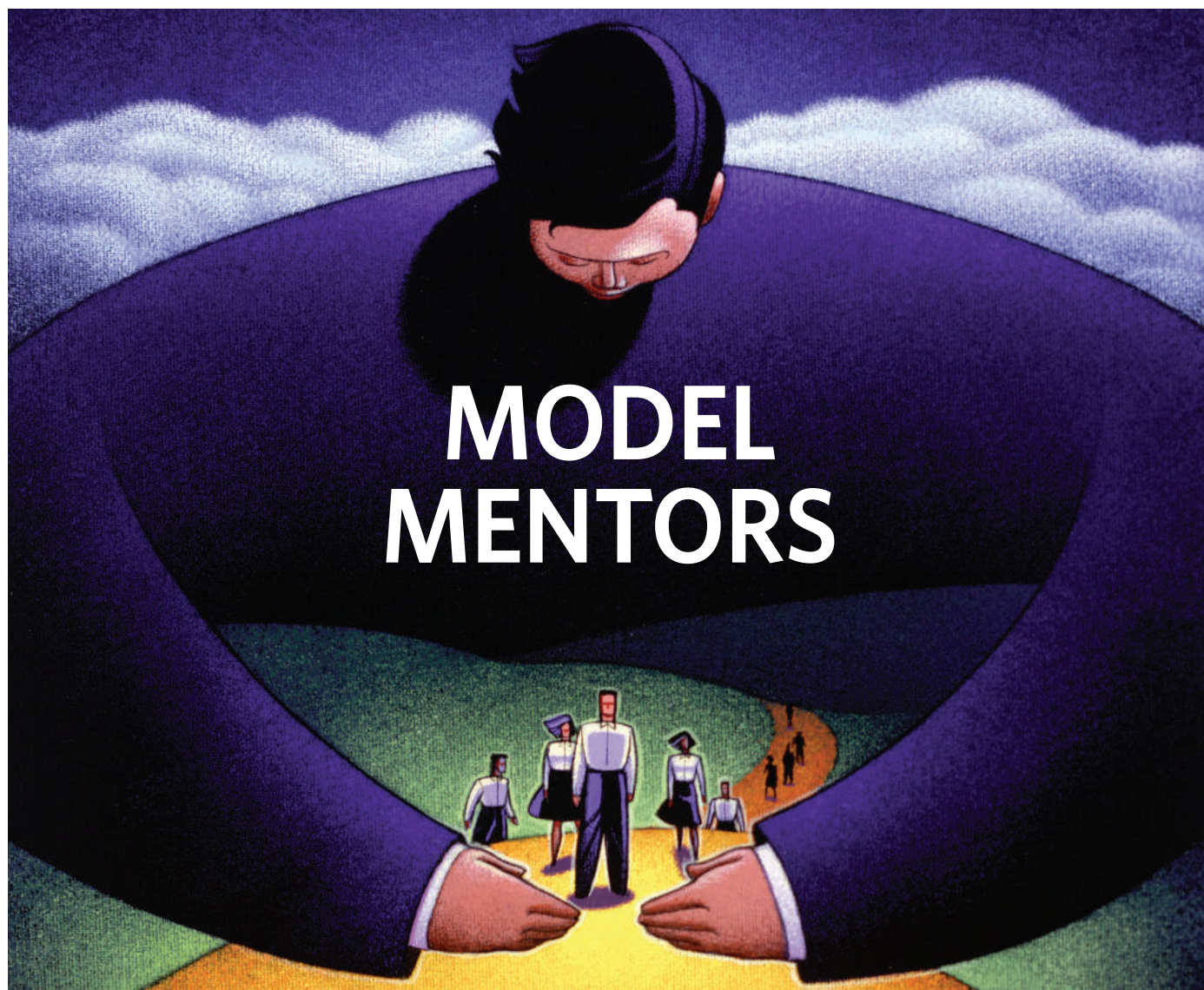
Naturejobs web development: Tom Hancock
Naturejobs online production:
Catherine Alexander

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com



D. TILLINGHAST/IMAGES.COM/CORBIS

MODEL MENTORS

Mentors are often called the 'unsung heroes' of science. A good mentor guides graduate students and postdocs out of an experimental quagmire or helps them take the long view when a project seems doomed. But their contribution to the scientific effort is rarely recognized.

This year, *Nature* has honoured five such heroes — three in Britain in partnership with NESTA, the National Endowment for Science, Technology and the Arts, and two in Australasia. In Britain, the lifetime achievement award was shared by evolutionary geneticist Godfrey Hewitt of the University of East Anglia and immunologist Andrew McMichael of the University of Oxford. Steve Watson, a British Heart Foundation researcher based at Birmingham University Medical School, won the mid-career award. Out of a field of 74 nominees from across Australia and New Zealand, Tom Healy, a physical chemist from the University of Melbourne, was selected to receive the lifetime award. Rachel Webster, an astrophysicist from the University of Melbourne, won the mid-career award.

What makes a good mentor? All five winners were recommended by current and former protégés, who attribute much of their own success to their supervisors'

Five scientists nominated by their peers have created nurturing research environments and fostered fields and careers far beyond their labs. **Carina Dennis and Janet Wright** give credit where it's long overdue.

guidance and encouragement. Scientific training was only part of it: these supervisors were prized for looking beyond short-term results, showing concern for individuals and helping them build sustainable collaborations beyond the mentors' purview. Their involvement ranged from helping with a recalcitrant experiment to organizing international conferences.

What they do not do is important too. Several nominators spoke about supervisors they were glad to have avoided — people who pit lab members against each other by giving them competing projects, who get angry if data don't support their own hypothesis, who are uninterested in their students, or who ignore any idea that they can't claim as their own.

A helping hand

Julia James marvelled at her mentor's hands-on approach. "In spite of his busy schedule, Professor McMichael donned a lab coat to assist me in growing a cell line," says James, who is working on her DPhil. "His level of commitment and interest in my success is inspiring."

Although he is hands-on when needed, McMichael also gives people space to develop their own ideas. "It's often a matter of helping them decide if something

they've found is a breakthrough or an artefact that they shouldn't waste time on," he says. "It's where experience helps a lot."

McMichael learned this approach from his own mentors, including Hugh McDevitt at Stanford, and has passed it on. One of his former protégés, Charles Bangham, now professor of immunology at Imperial College London, says: "I try to mentor people as Andrew did, being ready to provide criticism when asked for it, being open to new ideas, not being prejudiced against an idea just because it doesn't fit with my current ideas or wishes."

Good mentoring can continue long after a student leaves the lab. "I've known Andrew since 1982 and I still ask his advice," says Bangham. One of McMichael's own mentors, Brigitte Askonas of Imperial College London, still helps students at his lab on regular visits, more than 30 years after she supervised his PhD.

Source of skills

Sometimes students and postdocs don't know when to seek advice — especially for skills beyond science. Watson tries to identify such needs and have those in his lab improve these skills. "Nearly everyone says they're a bad speaker, or can't write research papers, but these are skills that can be taught," he says. "Clearly not everybody is successful at all aspects of research, however, so this needs to be more carefully tailored."

Mentors need to know when and how much to intervene. "It's important to give students and postdocs the freedom to make their own mistakes and follow their own ideas, but only within reason," Watson says. Letting young scientists know they can turn to their adviser for help offers reassurance, and can motivate them to do better work. "For the individuals in the lab, it really does matter that they receive the support they need, as this is their only chance for this stage of their career," says Watson. "If each individual receives the appropriate level of support — including a sense of fun, objective advice and independence — the research will follow."

One of Watson's former protégés, Alastair Poole, notes that offering such support requires self-confidence. "I think Steve naturally realizes that supporting others through their careers doesn't threaten your own career, but rather the opposite," says Poole, now a reader in pharmacology at the University of Bristol. He adds that providing such support can elevate not just a lab, but an entire field: Watson's support for younger scientists has helped raise the profile of UK platelet and thrombosis research.

Hewitt encourages young scientists to build research teams among themselves. He starts by choosing people who personally click within his lab, then encourages



"People who are very personally ambitious miss the point: they don't have the combined effects of lots of people working together."
— Godfrey Hewitt

them to build partnerships. "These days, with many skills impinging on a particular topic, you can work better, bounce ideas off each other, share the chores," he says.

Once the team is up and running, Hewitt provides them with a platform to shine. "I remember how Godfrey would drag the most senior scientists relevant to my research into the auditorium for my first talks at conferences, and then would nod encouragingly from the audience as I described my work," says Douda Bensasson, one of Hewitt's many former PhD students. Now a postdoc at the University of Manchester, Bensasson feels fortunate to have been advised to choose her graduate school for the supervisor she would have, rather than for a project or a place.

"There are some sharks around," says Hewitt. "But people who are very aggressive and personally ambitious miss the point: they don't influence others to do things, they don't have the combined effects of lots of people working together."

The team-building lesson has endured among his protégés. "Godfrey maintained that you can't expect to become an expert on everything, so it's important to build partnerships with people who can complement your own expertise," says Steve Cooper, now director of the Evolutionary Biology Unit at the South Australian Museum in Adelaide. He believes the collaborations he has built with people from different scientific fields have played an important role in his own research success.

Hewitt's team-building has repercussions long after his postdocs and students leave his lab. As students move on, they form an ever-growing network of contacts and expertise. And now that scientists he mentored have gone on to mentor their own students around the world, Hewitt takes an interest in the international careers of his 'academic grandchildren'.

Team effort

In Australia, Rachel Webster used a similar philosophy to create a strong research environment from scratch; there was very little astrophysics research at her institution when she was appointed just over ten years ago. Since then, she has fostered a thriving astrophysics community and spawned a significant pedigree of protégés, who attribute their success to her guidance.

Nominators emphasize Webster's wisdom in matching the individual with opportunities. "Webster sees the skills of each individual and then designs projects to suit these strengths," says Alicia Oshlack of the Walter and Eliza Hall Institute of Medical Research in Melbourne. She is commended not only for her continual attention to the career development of her protégés, but also for appreciating that not everyone will have the same career trajectory as hers.

"It really does matter that people receive the support they need, as this is their only chance for this stage of their career."
— Steve Watson





Rachel Webster and Tom Healy received their mentor awards at the Melbourne Planetarium (left).

"She assumes you are a complex person who also happens to be a scientist, instead of a scientist who it so happens turns out to be a complex individual," says Maurizio Toscano of the University of Melbourne. She also encourages a comprehensive outlook by emphasizing that even theoretical physicists need familiarity with concrete data.

"Webster argued that anyone planning on becoming a theorist ought to spend some time confronting the data," says David Hogg of New York University, who was first mentored by Webster as an undergraduate in the early 1990s. "So she put me on a project designed to make a young and naive proto-theorist understand that real data are complicated, dirty and stubborn." But the lesson was an inspiration rather than a deterrent — Hogg pursued more observational projects with large data sets and now runs a group dedicated to statistical astronomy.

Beyond the lab

Webster has also been a role model for women in physics. Ten years ago, she introduced the Women in Physics programme at the University of Melbourne; it has markedly bolstered the numbers of female graduates. "Webster was someone I felt I could relate to," says Annette Ferguson. "Until that point, I had almost exclusively encountered middle-aged male professors during my physics studies and most of my fellow classmates were also male." Ferguson, who first met Webster during her undergraduate years, is now a lecturer at the University of Edinburgh, UK.

Healy also helped build science beyond his own lab. "Tom Healy is an example of a person who not only mentors his own students, he also succeeds in mentoring an entire field of science," says William Ducker of the University of Melbourne. Australian colloid and surface science is ranked among the best in

the world, thanks in large part to Healy, Ducker says.

Healy brings people together from multiple fields to work on projects — whether or not they directly benefit his lab, says a former PhD student, Russell Crawford.

"He encourages research collaborations among scientists, even where he, or his university, would not be directly involved nor be beneficiaries of the research," says Crawford, now dean of the faculty of life and social sciences at Swinburne University of Technology in Hawthorn, Victoria.

A meeting of minds

One of Healy's most substantial legacies is a student conference on colloid and surface science. He and Bob Hunter of the University of Sydney established it nearly 40 years ago to help young researchers network with their peers. The conference has since been emulated around the world. "I can name many excellent Japanese scientists who benefited from research collaboration and personal association with the Healy academic family," says Toyoki Kunitake of the University of Kitakyushu, Japan.

Healy fosters a nurturing environment and creates opportunities for his students, such as encouraging them to get overseas experience. An award bearing his name was established at the University of Melbourne to fund student travel to an overseas conference or research centre. He is also credited with bringing back to Australia some of the finest minds in the field.

He has guided people across academia, government laboratories and industry, extending his mentoring beyond traditional boundaries. "In today's world, Tom's interest and enthusiasm for forging strong collaborative links with industry might be seen as nothing unusual, but in the 1960s and 1970s it was ground-breaking," says Brian Kavanagh, of the Water Corporation of Western Australia, who was mentored by Healy during his undergraduate studies and throughout his career.

Despite the hours they've spent on other people's careers, none of these mentors feels they've missed out. This could be because they're prodigiously hard-working: nominators noted the colossal energy they put into both research and mentoring. Awardees all agree that a harmonious atmosphere aids a lab's productivity. And all have produced scientists who are following in their tracks by keeping good mentoring alive. In that regard, the songs of these 'unsung heroes' will reverberate long after they retire from research. ■

Carina Dennis is Nature's Australasian correspondent. Janet Wright is a freelance writer based in London.



"Experience is useful in helping people decide whether something is a breakthrough or an artefact that they shouldn't waste time on." — Andrew McMichael

The Godmother Protocols

More reliable than kissing.

Heather M. Whitney

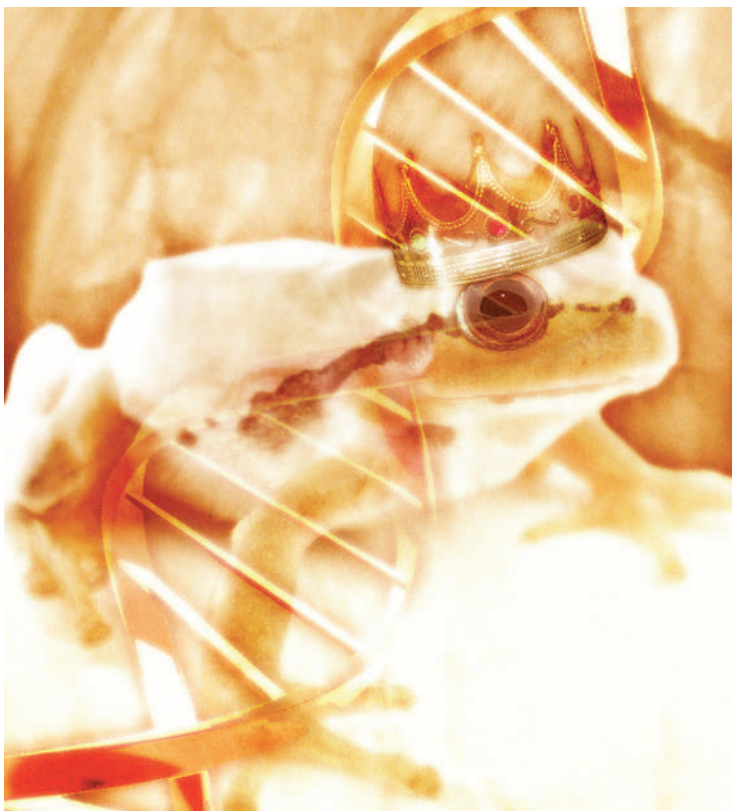
First take a frog. Choose your species with care. Although the skins of some frog species produce antimicrobials, and the ability to touch for the King's Evil is a splendid certificate of true royalty, other species have remarkably poisonous skins. A prince that causes death or mass hallucinations every time he shakes hands on a royal walkabout could result in an unfortunate revolution and republicanism.

Second, put any unhygienic thoughts of kissing out of your mind. Transfection plasmids designed to transform frog to human are far too unstable to be transmitted orally, and additionally have a distressing tendency to insert randomly into the genome, resulting in a cancerous mess rather than the required princely form. An adapted version of the old-fashioned method is still considered best in this case.

Obtain DNA from a suitable candidate. Please be careful during this step — the offspring of many royal families have bodyguards that are a little short-tempered with anyone approaching their charges with needles or syringes. It is also a bad idea to stalk the selected donor — a restraining order has never helped anyone. It is usually best to select a royal family with a number of obscure offshoots, that was deposed at some point during the twentieth century, or with a tendency for extra-pair copulation.

Remove oocytes from your selected frog. Several hundred will be required, the process still being somewhat unrefined; however, eye of newt and wing of bat are not needed, whatever the older textbooks say. As normal somatic-cell nuclear transfer methods have yet to produce viable embryos in frogs, a Transmogrifier™ kit should be used. Fortunately, this technology has recently been adapted to be used with microinjection equipment. One potential disadvantage of this method is

that significant amounts of the original oocyte DNA can recombine with your candidate DNA. Do not worry about this too much — royal offspring with a fondness for flies can do little harm. If the Emperor of Japan can be an expert in guppy taxonomy and the occupier of the throne of Great Britain and the Commonwealth talks to plants, a royal entomologist is not going to raise any eyebrows.



Incubate the cells as usual, then implant them into a surrogate mother. With an amphibian-mammal cross such as this, it is generally considered that a marsupial surrogate is the best compromise, particularly considering the superb nutritional quality of the milk provided once your clutch of princes has hatched.

Once you are assured of at least one healthy male child, social exposure becomes of primary importance. You must send the hatchlings to a school, the older the institution the better. Although it may appear irrational to put a not-entirely-human creature through a further de-humanizing experience, it does ensure that any of his remaining quirks and foibles that have not been crushed out can be laid squarely at the door of his educational

facility. The ability to ride a horse and to dance are the two most common abilities required by ruling families, so concentrate on these. You will probably have to pay extra for this tuition. Fortunately the ability to slay dragons is no longer required. All dragons hatched to date have, by EU legislation, been toothless, flameless and survive solely on blood-agar soup. This is probably for the best as the training for

this skill is notorious for reducing the number of viable princes available.

Once the growth period and education are over, you can now habituate your specimen to its natural habitat. This is where choosing the DNA of an obscure royal family is of benefit, particularly if they have a tendency to extra-pair copulations. The most basic of unsubstantiated stories will be more than enough, when supported with your specimen's DNA fingerprint and phenotype (a good strong phenotype, such as the Bourbon nose or the Hapsburg lip, is also a good idea).

Unfortunately by this stage, unless you have modified your own telomeres, you will be far too old to be a suitable mate for your prince, as it is well known that princes require mates at least ten

years their junior in virtually all societies. If you have planned for this, you will have imprinted your prince at an early age with either a scent or sound that will require him to marry a woman of your choosing, preferably your daughter for kin-selection and socio-political control purposes.

However, be careful as this leaves you in a rather precarious position — you will be the type of mother-in-law that has the power to change frogs into princes or princes into frogs. They have a tendency to come to a bad end. Make sure to keep your distance from open oven doors, ducking stools and stakes in the centre of pyres of wood that will inevitably appear whenever you visit.

Heather M. Whitney is a plant scientist living in Cambridge, UK.

JACEY